

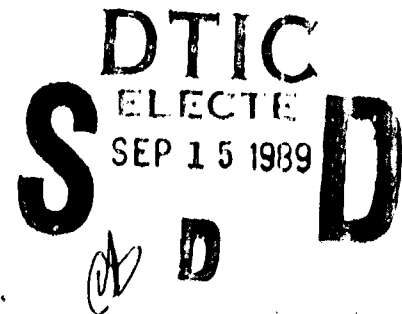
AD-A212 387

DTIC FILE COPY

**GaAs Gate Dynamic
Memory Technology**

**Final Report for
ONR Contract N00014-86-K-0350**

**Michael R. Melloch
James A. Cooper, Jr.
Thomas E. Dungan
Philip G. Neudeck
John W. Pabst**



**TR-EE 89-47
August 1989**

DISTRIBUTION STATEMENT A

**Approved for public release
Distribution Unlimited**

**School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907**

89

8

21

9

1

**GaAs Gate Dynamic
Memory Technology**

**Final Report for
ONR Contract N00014-86-K-0350**

**Michael R. Melloch
James A. Cooper, Jr.
Thomas E. Dungan
Philip G. Neudeck
John W. Pabst**

**TR-EE 89-47
August 1989**

**School of Electrical Engineering
Purdue University
West Lafayette, Indiana 47907**

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified			1b RESTRICTIVE MARKINGS	
2a SECURITY CLASSIFICATION AUTHORITY N/A since unclassified			3 DISTRIBUTION/AVAILABILITY OF REPORT	
2b DECLASSIFICATION/DOWNGRADING SCHEDULE N/A since unclassified				
4. PERFORMING ORGANIZATION REPORT NUMBER(S) TR-EE 89-47			5 MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Purdue University		6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Office of Naval Research	
6c. ADDRESS (City, State, and ZIP Code) School of Electrical Engineering West Lafayette, IN 47907			7b ADDRESS (City, State, and ZIP Code) Arlington, VA 22217	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research		8b OFFICE SYMBOL (If applicable) Code 1114	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0350	
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217			10 SOURCE OF FUNDING NUMBERS	
			PROGRAM ELEMENT NO.	PROJECT NO.
11. TITLE (Include Security Classification) GaAs Gate Dynamic Memory Technology				
12. PERSONAL AUTHOR(S) M.R. Melloch, J.A. Cooper, Jr., T.E. Dungan, P.G. Neudeck, and J.W. Pabst				
13a. TYPE OF REPORT Final Report		13b. TIME COVERED FROM 860601 TO 890531		14. DATE OF REPORT (Year, Month, Day) 890801
15. PAGE COUNT 148				
16. SUPPLEMENTARY NOTATION				
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB-GROUP		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) <p>The expected performance characteristics of GaAs dynamic memories are compared with the capabilities of existing technologies to establish a speed-capacity window for possible applications. The design of GaAs dynamic memories using FET direct-access of PN-junction-based storage capacitors is developed. The leakage mechanisms in PN-junction capacitors are considered theoretically, and experimental performance of mesa-isolated capacitors in GaAs and AlGaAs is reported. Optimization of storage-time performance and charge capacity by selection of materials and dopings is discussed, and the limitations of optimized capacitors with respect to temperature and scaling are examined experimentally. MBE-grown mesa-isolated PN-junction capacitors are demonstrated to have both sufficient storage time and sufficient capacity for high-density GaAs DRAMs operating above 100°C. Design of access transistors for optimal subthreshold performance is discussed. A two-dimensional harmonic solution for the potential in subthreshold FETs is presented. The har-</p> <p style="text-align: right;">(CONTINUED)</p>				
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a NAME OF RESPONSIBLE INDIVIDUAL Max Yodar			22b TELEPHONE (Include Area Code) 202-696-4218	22c OFFICE SYMBOL ONR, Code 1114

6. Abstract (continued)

monic solution is used to calculate the relationships between physical FET design parameters and subthreshold performance. Trade-offs between design for best subthreshold characteristics versus manufacturability and circuit requirements are considered. The design of complete DRAM cells combining a capacitor and an access transistor is developed. Required operating voltages for read, write, and storage sequences are established. The advantages and disadvantages of candidate cell configurations and fabrication techniques are discussed. Experimental demonstration of complete GaAs dynamic memory cells operating at well above room temperature is presented.

The research supported by this contract has resulted in 14 journal publications and 7 conference presentations.



Accession For	
NTIS	CRA&I <input checked="" type="checkbox"/>
DTIC	TAB <input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By <i>per ltr.</i>	
Distribution	
Availability Codes	
Dist	Accession and/or special
A-1	

TABLE OF CONTENTS

	Page
CHAPTER 1 - THE CASE FOR A GALLIUM ARSENIDE DYNAMIC MEMORY	1
1.1 Memory Performance Criteria.....	1
1.2 Integrated-Circuit Memory Implementation Technologies	3
1.3 Semiconductor Materials: Si Versus GaAs.....	3
1.4 Static Versus Dynamic Design.....	5
1.5 Reported Performance of Memory Implementation Technologies.....	8
1.6 Projected Performance of GaAs DRAMs.....	10
CHAPTER 2 - GaAs STORAGE CAPACITOR IMPLEMENTATION.....	13
2.1 Minority-Carrier Devices	13
2.2 PN-Junction Capacitors	15
2.2.1 Charge-Storage Density	16
2.2.2 Storage Time	19
2.2.2.1 Bulk Depletion Current	21
2.2.2.2 Bulk Diffusion Current.....	22
2.2.2.3 Perimeter Depletion Current	25
2.2.2.4 Perimeter Diffusion Current.....	32
2.2.2.5 Bulk and Perimeter Limited Storage Time Expressions	39
2.2.3 Experimental Demonstration of PN-Junction Capacitors	40
2.2.4 Details of the Capacitance-Transient Measurement Technique	48
2.2.4.1 Initial-Bias Dependence of the Storage Time	48
2.2.4.1.1 Relation of Capacitance-Recovery Time Constant to Charge-Recovery Time Constant	49
2.2.4.1.2 Non-Exponential Charge Recovery in Asymmetric Junctions	53
2.2.4.2 Charge Redistribution During the Bias Pulse.....	57
2.2.5 Experimental Capacitor Performance Versus Doping Level	60
2.2.5.1 Field-Enhanced Generation.....	63

	Page
2.2.5.2 The P ⁺ iN ⁺ Capacitor.....	67
2.2.6 Storage Time Dependence on Perimeter-to-Area Ratio.....	70
2.2.7 Ion-Implanted Capacitors.....	72
2.2.7.1 Mg and Proton Implant-Isolated Devices.....	74
2.2.7.2 Capacitors Formed by Si and Mg Implantation	76
CHAPTER 3 - ACCESS TRANSISTORS.....	78
3.1 Introduction	78
3.2 Subthreshold Leakage	80
3.2.1 Harmonic Solution for Potential in Subthreshold FETs.....	83
3.2.2 Parameter Extraction from 2-D Potential Solutions.....	91
3.2.2.1 Calculation of Subthreshold Ideality Factor and Subthreshold Swing	92
3.2.2.2 Calculation of Threshold Voltage	94
3.2.3 Calculated Dependences of Subthreshold Parameters on Dimensions and Dopings.....	94
3.2.4 Selection of Design Parameters for Optimum Subthreshold Performance.....	104
3.3 Measured FET Subthreshold Performance.....	106
CHAPTER 4 - COMPLETE DYNAMIC MEMORY CELLS	112
4.1 Operating Voltages.....	112
4.2 Cell Configurations.....	116
4.3 Experimental Results from Complete Cells	121
CHAPTER 5 - RECOMMENDATIONS FOR FUTURE RESEARCH.....	133
5.1 Introduction	133
5.2 Improving Experimental Cells	133
5.3 Continued Ion-Implantation Research	136
5.4 Experimental Investigation of Transistor Scaling Effects	138
5.5 Circuit Development	138
LIST OF REFERENCES	140
APPENDIX A - JOURNAL PUBLICATIONS AND CONFERENCE PRESENTATIONS ACKNOWLEDGING SUPPORT FROM CONTRACT N00014-86-K-0350 AND THESES.....	146

CHAPTER 1

THE CASE FOR A GALLIUM ARSENIDE DYNAMIC MEMORY

1.1 Memory Performance Criteria

Integrated-circuit semiconductor memories are desirable for non-permanent digital information storage because of their high speed, low cost, compact size, and compatibility with integrated-circuit central processing units. In the design of any digital system, the technology used to implement the system is selected by determining which available technology can provide the demanded system performance at the lowest cost. System performance parameters include operating speed, power dissipation, physical size and weight, and reliability in the specified environment. In some applications, performance requirements have well-defined upper and lower limits. For example, there is no need for digital controllers in home appliances to operate at several MHz, and the weight and power requirements of a land-based main-frame computer are of little consequence. In many applications, however, the requirements are more open-ended. Super computers, for example, can never be fast enough, and increasing the reliability of any system is always desirable. In such situations, the optimal implementation technology is determined by evaluating the trade-offs between improving performance of an open-ended criterion and system cost, or between competing criteria such as speed and power.

Specific performance characteristics differentiating semiconductor memory technologies are access time, power-dissipation-per-bit, single-chip capacity, operating temperature range, radiation hardness, and cost-per-bit. Some of these specific characteristics are in one-to-one correspondence with the more general system performance requirements, while others are interrelated. For example, power-dissipation-per-bit and cost-per-bit translate directly into system power and system cost, while system reliability is affected by operating temperature range, radiation hardness, and single-chip capacity. Chip capacity is related to reliability because higher capacity per chip reduces the number of pins and chip-to-chip interconnects, which can be significant sources of system failure.

The relationship between the specific characteristics of memory technologies and the operating speeds of systems in which the technologies are used can become complex, particularly in systems designed for very high throughput. This is because memory-system speed performance is a strong function not only of the specific characteristics of the individual integrated circuits employed, but also of the architecture used to define their functions. For example, memory systems which must handle large amounts of data at high speeds are designed hierarchically, with a limited amount of very fast, high-cost memory backed by a larger and slower main memory implemented in a less expensive technology. The small, fast memory block, called the cache, would ideally be located on the chip with the processor itself. This may be impossible due to incompatible processor and memory technologies or because of power-dissipation or chip-capacity restrictions. If the cache is on a separate chip or chips, it must be located as close as possible to the processor to minimize interconnect delays. A memory technology with very limited single-chip capacity could require many chips to compose the cache, making a compact arrangement difficult, and possibly reducing performance.

When evaluating the impact of specific characteristics of a memory implementation technology on system speed performance, it is essential to identify the level of the memory hierarchy where the technology would be employed. For example, a technology with a relatively slow access time but with very low power-dissipation-per-bit and high single-chip capacity could be used as main memory in a cached system without significantly limiting speed performance. As a second example, consider a memory technology intended for use in a high-speed cache not located on-chip with the processor. In this case, it is critical that the single-chip capacity be great enough to allow the cache to be implemented in few enough chips to be placed in close proximity to the processor. Once this goal is achieved, there is a sharply diminishing return in system speed performance for increasing single-chip capacity, because increasing cache size beyond a certain point is ineffective.

Consideration of hierarchical memory systems also points out another less-easily-quantifiable but possibly very important characteristic of memory implementation technologies: the ease with which one technology can be interfaced with others. For example, it may be desirable to use different technologies with different electrical input-output characteristics for main memory and a cache. A more extreme case is in the design of processors with on-chip cache, where the memory technology and the processor must have not only compatible electrical characteristics but compatible fabrication sequences as well.

1.2 Integrated-Circuit Memory Implementation Technologies

Broadly defined, an integrated-circuit memory technology is a choice of a semiconductor material, a dynamic or static approach, and a logic form and transistor type. For example, the dominant form of super-computer cache memory is currently silicon static emitter-coupled logic (ECL), which uses bipolar transistors. Very high density, low cost memory is implemented using silicon dynamic designs in complementary metal-oxide-semiconductor (CMOS) logic. Such classifications of memory technologies are becoming less distinct, however, due to efforts to combine the best features of different approaches. Recently designs have been demonstrated which use different logic forms for the memory array and the peripheral access circuitry, such as silicon BiCMOS, which uses bipolar peripheral circuitry with a CMOS static memory cell array [1]. Development is even underway for single integrated circuits which have components implemented in separate semiconductor materials, such as GaAs on Si [2].

Within a broad classification, such as silicon static CMOS, there are an unlimited number of design variations and details which will affect the specific performance characteristics of the memory. For example, access time and single-chip capacity may be strongly dependent on minimum feature size. The performance characteristics of all technologies tend to improve as the designs gain maturity. These considerations make it difficult to rigidly specify the performance limitations of any particular technology. However, it is possible to identify the types of applications for which a certain technology is best suited by examining its demonstrated and projected performance characteristics.

The task of determining the usefulness of GaAs dynamic memory consists, then, of identifying the potential applications for which it would be best suited, considering the competing technologies, and comparing the relative strengths of the projected performance characteristics.

1.3 Semiconductor Materials: Si Versus GaAs

The choice of semiconductor materials currently capable of supporting large-scale integrated circuits is limited. There are two elemental group IV semiconductors, Ge and Si, and a large number of compound semiconductors consisting of elements from groups III and V, e.g. GaAs, or of elements from groups II and VI, e.g. ZnSe. Of the elemental semiconductors, silicon's electronic properties and natural processing advantages make it obviously better suited for large-scale integration. Of the compound materials, the only system

currently well-enough developed and researched for commercial production of digital integrated circuits is GaAs.

Silicon is the dominant semiconductor used for integrated circuits in general, and for memories in particular, because of its mature fabrication procedures and the low cost of very-high-purity material. GaAs wafers are currently about twenty-five times as expensive as silicon per-unit-area, and the size of available wafers is limited to a four-inch diameter as compared with six- or even eight-inch wafers in Si. Defect densities tend to be higher in GaAs than in Si due to the inherent difficulties in producing the more complicated crystal-lattice structure of GaAs. In addition, GaAs is a poorer thermal conductor than Si, so removing heat produced by high-density circuits is more difficult. Higher defect densities and poorer thermal conduction combine to limit die sizes and transistor counts in GaAs to levels well below those used in some Si designs [3]. Finally, the lack of a useful native oxide of GaAs complicates both fabrication procedures and device designs.

GaAs does have some advantages over Si which might make it the material of choice in situations with extreme environmental or performance requirements. GaAs has a wider bandgap than silicon, which can make GaAs circuits more tolerant of temperature extremes. GaAs circuits are, in general, less susceptible to radiation damage, largely due to the absence of devices dependent upon oxides [4]. These two factors make GaAs an attractive candidate for aerospace and military applications.

Perhaps more importantly, the band structure of GaAs results in an electronic effective mass which is about fifteen times less than in Si. The same electric field accelerates electrons faster in GaAs than it would in Si, allowing transistors to switch faster for the same power consumption. Another speed advantage of GaAs comes from the availability of semi-insulating GaAs substrates [5], which reduce transistor parasitic capacitances and interconnect capacitances. Maturing GaAs circuit technologies have entered the VLSI regime and offer the possibility of producing microprocessors which could run at clock rates five times that of the best current Si designs [6].

Another feature which could eventually provide an even larger speed advantage for GaAs over Si comes from the fact that GaAs is nearly perfectly lattice-matched to a wider-bandgap III-V compound, AlAs. This allows epitaxial formation of atomically-abrupt hetero-interfaces between GaAs and AlAs or between GaAs and the ternary compound $\text{Al}_x\text{Ga}_{1-x}\text{As}$. Heterojunction-based field-effect transistors have demonstrated even greater speed-power performance than conventional metal-gate GaAs FETs [7]. Heterojunctions also make high-

gain bipolar transistors available in GaAs [8], presenting the possibility of GaAs imitations of silicon's ECL and BiCMOS.

With the direct bandgap of GaAs, the GaAs-AlAs material system is also an excellent choice for fabrication of light-emitting diodes and lasers. As speeds of digital systems continue to increase, chip-to-chip delays are becoming a more serious limitation. Also, the number of pins required by complex circuits is continually increasing. In the future, chip-to-chip communication may be accomplished by modulated laser emission over optical fibers. Time multiplexed fibers could be used to reduce the number of I/O lines required without decreasing performance.

Finally, many of the undesirable properties of GaAs may be eliminated by new mismatched epitaxial growth techniques which allow growth of thin, relatively-high-quality GaAs layers on Si substrates [9]. This approach allows fabrication of larger, more economical GaAs films, and provides the heat-sinking capabilities of Si. GaAs-on-Si has shown rapid advances, but some difficulties still remain. Stresses due to the different thermal expansion coefficients of the two materials limit the GaAs layer thickness, resulting in higher parasitic capacitances than is found with pure GaAs. The stress levels currently experienced are also too high for fabrication of long-lifetime lasers.

1.4 Static Versus Dynamic Design

A standard six-transistor static memory cell uses two cross-coupled inverters to form a bistable latch, as shown in Figure 1.1 [10]. One stable state corresponds to a logic "1" and the other to a logic "0". The cell is static in that, in the absence of an intentional "write" sequence, it will remain in either of its stable states as long as the power supply is uninterrupted.

A dynamic design achieves higher integration density by reducing the number of cell elements from six to two. A standard one-transistor dynamic memory cell, shown in Figure 1.2, consists of an access transistor and a capacitor. A logic "1" or a logic "0" is represented by the presence or absence of charge on the capacitor. Because leakage currents will cause the stored charge to disappear over time, a dynamic cell must be periodically "refreshed" by reading and rewriting the device. To operate efficiently, the leakage rate must be small enough so that the refresh rate is very slow compared to the operating frequency of the dynamic memory.

Information is written to the dynamic cell by pulling the bit line either high or low and turning on the access transistor momentarily to charge or discharge

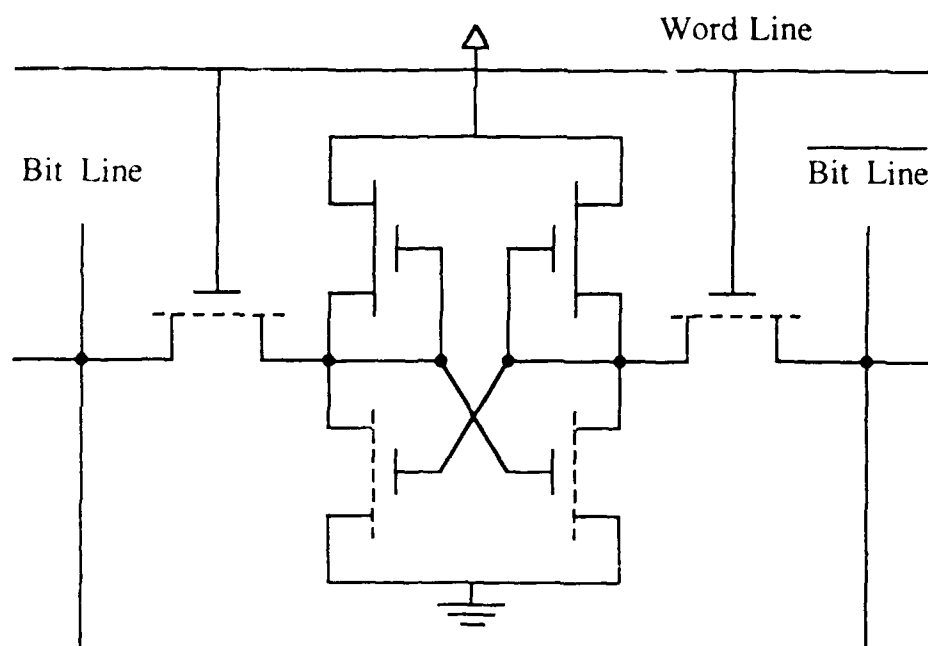


Figure 1.1 A six-transistor enhancement-depletion static memory cell.

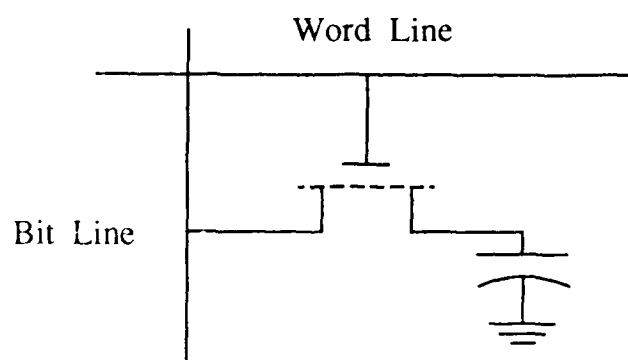


Figure 1.2 A one-transistor dynamic memory cell.

the cell capacitor. To write the static cell, either the true or complemented bit line is pulled low while the other is held at an intermediate value. The access transistors are then turned on, and current flows through the load device and access transistor connected to the low bit line. The ratio of access transistor "on" conductance to load device "on" conductance must be made large enough so that the gate of the opposite driver will be brought below threshold, setting the state of the cell.

To read the dynamic cell, the bit line is precharged to a voltage value half way between the voltage of a stored "0" and a stored "1". The access transistor is then turned on, so that capacitive charge sharing occurs between the cell capacitor and the bit line. If the small cell capacitor is at a low voltage, the bit-line voltage will decrease slightly; if the cell capacitor is at a high voltage, the bit-line voltage will increase slightly. "Sense-application" circuitry must be provided for each bit line (each column of the array) to restore the small bit-line voltage swing to the full voltage levels required to distinguish a logic "1" or a logic "0".

The static memory cell could, in principle, pull the two complementary bit lines all the way to the full logic-level voltages during a read cycle without sense amplification. In practice, however, sense amplifiers are also used in static designs to maintain speed performance. This is required because the cell's ability to source current is limited due to the necessarily small "on" conductance of the load devices. In theory, only a single bit line is needed to operate the static cell, but in practice reliable full-speed operation requires the two complementary lines.

Both static and dynamic memories can suffer "single-event upsets" or "soft errors" due to ionizing radiation. In the case of the dynamic cell, free carriers created by ionizing radiation can be collected directly by the cell's storage capacitor, upsetting the stored charge. In the static cell, carriers can be collected by the parasitic capacitance of the driver transistor's gate, possibly causing switching of the cell's state. In both static and dynamic designs, soft errors can occur because of radiation striking bit lines or sense amplifiers. Several techniques have been suggested and employed to reduce radiation-induced soft-error rates in both dynamic and static memories. It is difficult to definitely identify whether static or dynamic approaches have an inherent advantage in soft-error immunity. It is well established, however, that soft-error rates increase in all designs as cell dimensions are reduced.

Both dynamic and static memory cells also have upper limits on performance with respect to temperature. Leakage currents in dynamic

memories are usually thermally activated; that is, the leakage rate increases exponentially with temperature. Thus if the refresh rate is held constant, at some temperature insufficient charge will remain in the storage capacitors at the end of a cycle, and the stored information will be lost. Temperature-induced failure mechanisms in static memories are more complicated, involving increasing subthreshold currents in the access transistors and drivers [11]. For both approaches, observed temperature limitations are largely dependent on design details such as transistor type choices, noise margins, and cell size. As with radiation-induced soft-error rates, it is not clear that static or dynamic designs are inherently superior with respect to operating temperature limitations.

Power dissipation-per-bit in static memories varies widely depending on the type of transistors chosen to implement the cell. The static cell shown in Figure 1.1 is implemented with enhancement and depletion transistors in direct-coupled FET logic (E/D DCFL). This design might be used in GaAs or silicon NMOS. One of the two drivers is always turned on, so current is continually being drawn from the power supply through the associated load device. Power dissipation can be reduced at the expense of switching speed by replacing the depletion load devices with very large resistors. Power dissipation in the storage state can be virtually eliminated by using complementary logic, that is, replacing the load devices with P-channel FETs. In a complementary cell, all six transistors are turned off during storage, and significant power dissipation occurs only when the cell changes state.

The dynamic cell can achieve the same low-power performance of complementary static cells without the need for a P-channel FET. If the access transistor is enhancement mode, then the word line is held at ground during the storage state, and the dynamic cell dissipates almost zero power. Even with a depletion-mode access device, the power dissipation of the dynamic cell comes only from the gate leakage of the access transistor. Dynamic memories can then be expected to have a large power dissipation advantage over static designs in situations where complementary logic is not available, such as in GaAs MESFET logic.

1.5 Reported Performance of Memory Implementation Technologies

Section 1.1 listed six specific performance characteristics for integrated-circuit memory technologies: operating temperature range, radiation hardness, access time, single-chip capacity, power-dissipation-per-bit, and cost-per-bit. The first two characteristics, operating temperature range and radiation

hardness, usually influence technology choices only when an unusually severe operating environment is expected. All of the technologies considered are capable of performing over the commercial operating temperature range of 0° to 85° C. All of the technologies are also assumed to be capable of withstanding normal ground-level radiation doses. In general, any technology can be redesigned at the device, circuit, and packaging levels to extend both its operating temperature range and its radiation hardness, possibly at the expense of a small reduction in its other performance characteristics. For these reasons, radiation hardness and operating temperature range will not be considered further, except to repeat the general statements made in Section 1.4. Technologies not dependent on oxides, such as MESFET and bipolar, have an intrinsic radiation hardness advantage over MOS, and wider-bandgap semiconductors have an intrinsic advantage in operating temperature range, all other factors being equal.

The four remaining performance characteristics are intimately related. Power-dissipation-per-bit and cost-per-bit decrease for any given technology as single-chip capacity increases. Similarly, access times tend to increase with single-chip capacity. Basically, integrated circuit memory applications can be divided into two areas: high performance and high efficiency. The high-performance area emphasizes minimum access time for very-high-speed computers, image processors, and other real time applications. The high-efficiency area demands minimum cost-per-bit for high-volume, non-speed-critical applications such as main memories. The minimum-access-time requirement means that the high-performance area occupies the low end of the single-chip capacity spectrum. The high-efficiency area is at the high end of the single-chip capacity spectrum where minimum cost-per-bit is achieved. Presently the dividing line between the two areas is roughly between 256 Kb and 1 Mb single-chip capacities.

The high-efficiency area is impenetrably dominated by silicon MOS designs. Silicon bipolar cannot achieve the ultra-high chip capacities because of power dissipation limitations, and even if it could, the added complexity of bipolar fabrication means that it cannot compete on a cost-per-bit basis with MOS. Similarly, GaAs static memories cannot produce extremely high single-chip capacities due to cell-size and power-dissipation limitations. Even if dynamic designs were able to overcome the cell-size and power-dissipation limitations, GaAs DRAMs cannot compete with their Si counterparts in the high-efficiency area because of the large difference in the cost of the materials. For these reasons, the rest of this discussion will focus only on the high-performance area.

Figure 1.3 gives a summary of the access-time and single-chip-capacity performances reported in the last few years for several different memory technologies, in the high-performance area. The vertical axis represents access time and the horizontal axis represents single-chip capacity, so the optimum performances are at the lower right area of the graph. The scatter in the data for any given technology represents the effects of continuing performance improvements as the technology matures. The figure indicates the general characteristics of the various implementation technologies. Bipolar Si ECL can provide very short access times at relatively limited capacity. Si static MOS is just beginning to enter the high-performance area at the upper limits of both single-chip capacity and access time. Si BiCMOS bridges the gap between all-bipolar ECL and all-MOS designs, providing fairly high single-chip capacities and fairly short access times. All of the GaAs memories pictured are, of course, static designs. In general they compete with Si ECL at the extreme high-speed, low-capacity end of the scale.

1.6 Projected Performance of GaAs DRAMs

To determine where dynamic memories in GaAs might appear on the speed-capacity plot of Figure 1.3, it is interesting to consider the relationship of static and dynamic memories in Si. The Si dynamic memories reported in the same years lie outside the range of the figure. They are in the high-efficiency area above and to the right of the Si SRAMs, having capacities of 4 and 16 Mb and access times greater than 50 ns. Similarly, GaAs DRAMs can be expected to have performances placing them to the right and slightly above the GaAs SRAM designs. However, competition from other technologies in the high-performance area limits the increase in access time that can be tolerated going from static to dynamic design in GaAs. A reasonable window for the targeted performance of GaAs DRAMs using current technology is then represented by the dashed region in Figure 1.3. The window extends from about 3 ns at 16 Kb to 5 ns at 256 Kb. It is an increasingly important performance region because very-high-speed processors are requiring larger and larger cache memories, which are awkward to construct with only 1 and 4 Kb capacity chips.

There is already a reported performance point for Si ECL within that window at 3 ns, 32 Kb, and others are not far away. Maturing GaAs static designs are also moving toward 16 and 64 Kb. It is apparent that GaAs DRAMs cannot depend on establishing a speed-capacity region which is not attainable by competing technologies, as is the case with Si DRAMs. For GaAs dynamic memories to find applications, they must have other distinguishing

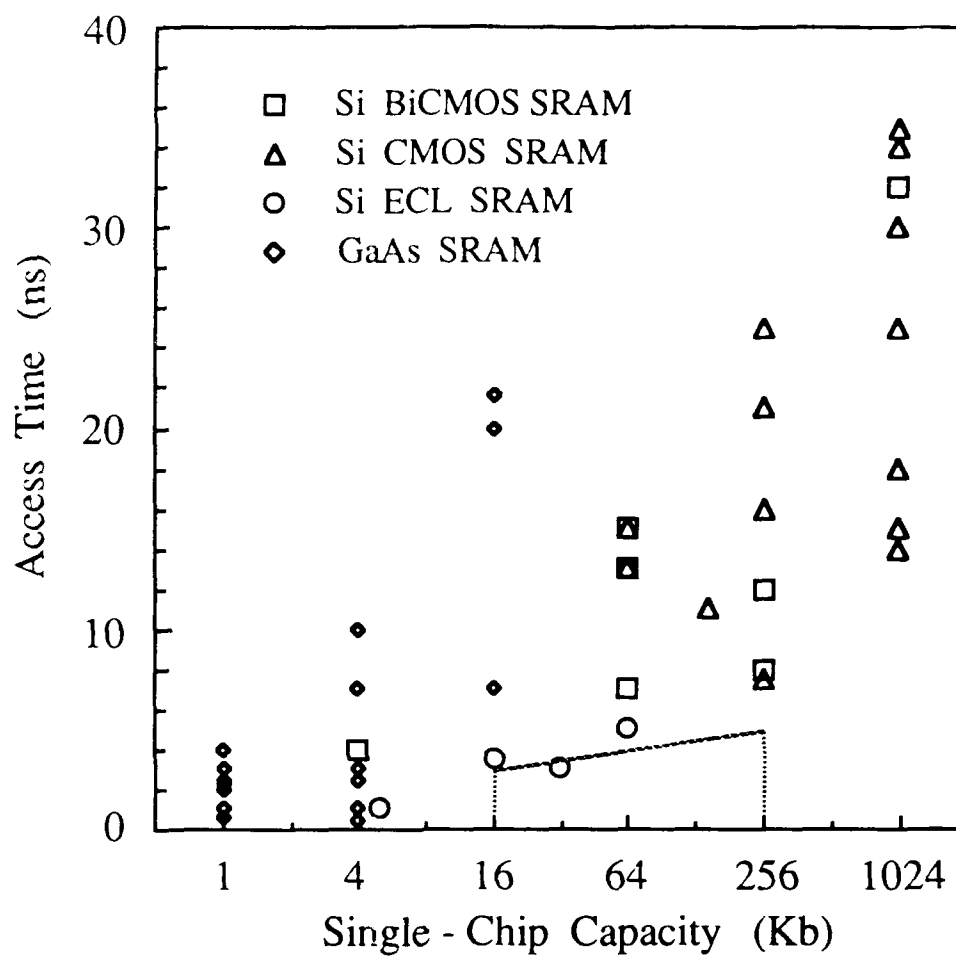


Figure 1.3 Survey of access-time versus single-chip capacity for high-performance SRAMs reported between 1985 and 1988.

performance characteristics which make them more attractive than their competitors, namely Si ECL and BiCMOS and GaAs static memories.

When comparing GaAs dynamic memories to static designs, one advantage of dynamic designs comes from the possibility of greatly reduced power-dissipation-per-bit in the storage state, as was mentioned at the end of Section 1.4. Complementary GaAs MESFET technology is not available, and enhancement-depletion static cells dissipate considerably more power than enhancement-accessed dynamic cells. Complementary JFET logic is available in GaAs, but it carries a large speed penalty, as is illustrated by the 16 Kb, 22 ns data point in Figure 1.3. Projected minimum cell sizes of dynamic and static designs are comparable, taking into account the continued shrinking of lithographic tolerances, so maximum single-chip capacity may be largely determined by power-dissipation limitations.

In comparing GaAs dynamic memory with all-bipolar Si SRAMs, the reduced power consumption of the GaAs DRAM is also an advantage. Si BiCMOS designs largely overcome the ECL power dissipation problems, however. The major advantage of GaAs DRAMs over Si BiCMOS designs is the natural compatibility of a GaAs DRAM with GaAs processors. GaAs-to-ECL interfaces are feasible [12], but they are an awkward detail which could be avoided by mating GaAs processors to GaAs memories, or by using Si ECL processors with Si bipolar memories. Whether or not this processor-compatibility factor actually provides an advantage for GaAs DRAMs depends strongly on the continued development of GaAs CPUs.

Finally, it should be realized that this is a relatively short-term analysis. The ultimate capabilities of GaAs DRAMs will be determined by design features which are currently being investigated. BiCMOS analogies in GaAs may significantly improve the performance characteristics of both GaAs SRAMs and DRAMs. Bipolar and MODFET access of GaAs dynamic memory cells are both being researched. Three-dimensional integration techniques may significantly reduce GaAs DRAM cell sizes, offering improvements in both access-time and single-chip capacity. The digital integrated-circuit technology picture changes very rapidly. Development of a dynamic memory for GaAs offers an additional option to the designers of future generations of ultra-high-performance digital systems.

CHAPTER 2

GaAs STORAGE CAPACITOR IMPLEMENTATION

2.1 Minority-Carrier Devices

Standard dynamic memory cell designs used in Si cannot be directly implemented in GaAs because they rely on the excellent electrical characteristics of the Si-SiO₂ interface. Most silicon dynamic memories are designed so that a logic "0" or a logic "1" is represented by the presence or absence of charge on an MOS capacitor, as indicated in Figure 2.1. The capacitor plate is held at a voltage which would result in the formation of a minority-carrier inversion layer in equilibrium. The equilibrium full-inversion-layer state corresponds to a logic "0". To produce a logic "1", the inversion layer is drained off through an access transistor, producing a non-equilibrium deep-depletion state in the capacitor. The storage time in such a design is determined by the rate at which generation of electron-hole pairs transforms the non-equilibrium logic "1" back into a logic "0". A very low interface-state density is essential to yield a generation rate low enough to allow useful storage times.

The difficulty encountered in trying to directly transfer such a design to GaAs is that the native oxides of GaAs are inadequate for use as the capacitor dielectric due to their poor qualities as insulators [13]. Other dielectrics such as SiO₂ or Si₃N₄ can be deposited on GaAs, but the resulting MIS structures do not exhibit the desired inversion/deep-depletion characteristics [14]. This is because the high density of surface states on the GaAs surface effectively pins the Fermi level at the surface, preventing modulation of the minority-carrier charge densities by the metal electrode [15]. In order to produce a MIS capacitor useful for minority-carrier-based dynamic memories in GaAs, a surface preparation technique capable of reducing surface-state densities would be required before deposition of the dielectric. Photochemical [16] and chemical treatments [17-18] have been used to alter the GaAs surface-state density, but the beneficial effects of these treatments disappear as the surface ages. A recent (NH₄)₂S passivation technique has been shown to increase the sensitivity of GaAs Schottky barrier heights to metal workfunctions [19], indicating at least

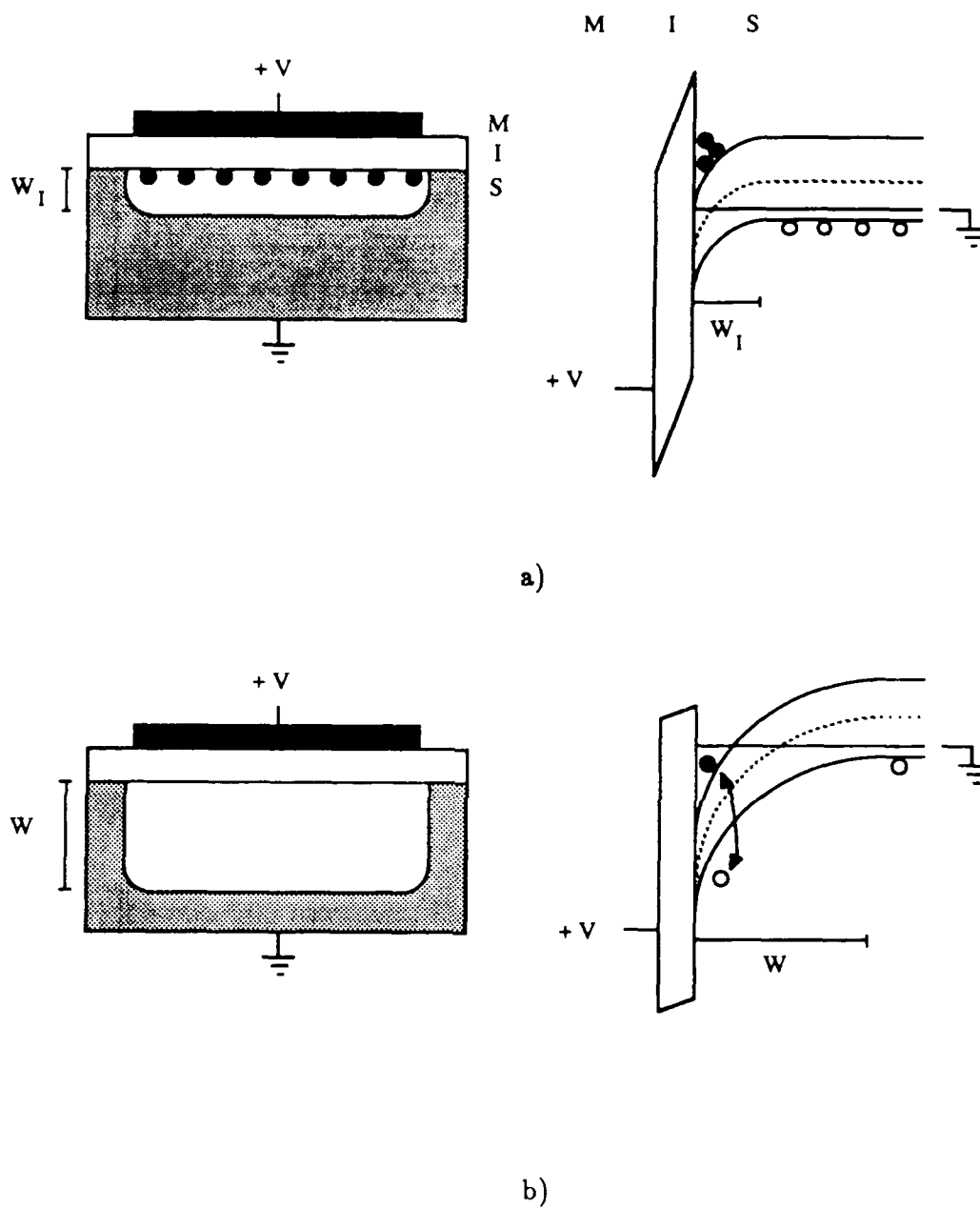


Figure 2.1 Logic "0" and logic "1" states of an MIS capacitor.

- a) Logic "0": equilibrium inversion
- b) Logic "1": non-equilibrium deep depletion

partial unpinning of the surface Fermi level. Although this $(\text{NH}_4)_2\text{S}$ treatment does not appear to suffer from aging effects when protected by a metal layer, calculation of the reduced surface-state density yields a number high enough to interfere with operation as a normal MIS capacitor, and compatibility with an available dielectric has not yet been demonstrated.

Another alternative is to replace the oxide with a lattice-matched wider-bandgap semiconductor such as AlGaAs or AlAs [20-22]. The interface-state densities at such heterojunctions are very low compared to the free-surface trap density [23]. Unfortunately, the electrostatic barrier confining the minority-carrier inversion layer at the heterojunction is very small compared to the Si-SiO₂ barrier, and thermionic emission limits the storage time. Such heterojunction-capacitor minority-carrier cells could provide a GaAs dynamic memory for liquid-nitrogen-temperature operation, but for room-temperature operation and above, another approach is needed. A GaAs floating-gate device using AlAs barriers has been reported with room-temperature storage times of a few seconds [24]. The explanation given for the long storage times is that the direct-to-indirect bandgap transition required for emission of electrons from GaAs to AlAs suppresses emission probabilities sufficiently to offset the low barrier height. This explanation is in direct conflict with subsequent experimental and theoretical demonstrations [25], indicating that the long storage times are probably due to carriers trapped at defect sites in the AlAs.

A current research topic is the use of slightly mismatched (pseudomorphic) materials as insulators for GaAs MIS structures. Pseudomorphic zinc selenide/N-GaAs capacitors have demonstrated accumulation-depletion-inversion behavior, but the small conduction-band discontinuity prevents electron inversion layer formation in ZnSe/P-GaAs samples [26]. Another approach uses a very thin pseudomorphic silicon layer to passivate the GaAs surface before deposition of SiO₂ as an insulator [27]. This technique has produced both electron and hole inversion layers, with a reported storage time of four minutes at room temperature for devices using holes as minority carriers. However, the capacitance-voltage characteristics show hysteresis corresponding to an interface-state density of $10^{12}/\text{cm}^2$, and attempts to fabricate inversion-mode transistors have so far been unsuccessful.

2.2 PN-Junction Capacitors

Because of the described difficulties in producing deep-depletion MIS capacitors in GaAs, other possible implementations of the dynamic memory storage capacitor must be considered. The two most critical requirements for

the capacitor in a dynamic memory are sufficiently long storage time over the intended range of operating temperatures and large capacity per-unit-area. One candidate solution which appears to satisfy both requirements is the replacement of the MIS capacitor with the capacitance of a PN junction. For example, consider a PN junction with the P region grounded and the N region floating, as indicated in Figure 2.2. The logic "0" state of the capacitor consists of zero voltage across the junction, with the depletion region at its equilibrium width. The non-equilibrium logic "1" state is produced by removing majority-carrier electrons from the floating N region, creating a reverse bias across the junction and widening the depletion region.

2.2.1 Charge-Storage Density

The charge stored on the PN-junction capacitor in the logic "1" state is defined as the ionized dopant sites uncovered on either side of the junction by the widening depletion region:

$$Q \equiv qN_A(x_p - x_{p0}) = qN_D(x_n - x_{n0}) \quad [2.1]$$

where Q is the charge storage density per-unit-area, q is the electronic charge, N_A and N_D are the acceptor and donor doping densities in cm^{-3} , and x_p , x_n , x_{p0} , and x_{n0} are depletion widths defined in Figure 2.3.

Applying the depletion approximation and Gauss's law to a uniformly-doped one-dimensional PN junction yields:

$$x_{n0} = \left(\frac{2\epsilon}{q} \frac{N_A}{N_D(N_A + N_D)} V_{bi} \right)^{1/2} \quad [2.2]$$

where ϵ is the dielectric constant and V_{bi} is the built-in potential of the junction. The expression for x_{p0} is obtained from equation 2.2 by interchanging N_A and N_D , and x_n and x_p are obtained from x_{n0} and x_{p0} by replacing V_{bi} by $V_{bi} + V_R$, where V_R is the applied reverse-bias voltage. For non-degenerately doped material,

$$V_{bi} = \frac{kT}{q} \left(\ln \left(\frac{N_D}{n_i} \right) + \ln \left(\frac{N_A}{n_i} \right) \right) \quad [2.3]$$

where k is Boltzmann's constant, T is absolute temperature, and n_i is the intrinsic carrier concentration.

The built-in potential can be expressed for both degenerately and non-degenerately doped material by:

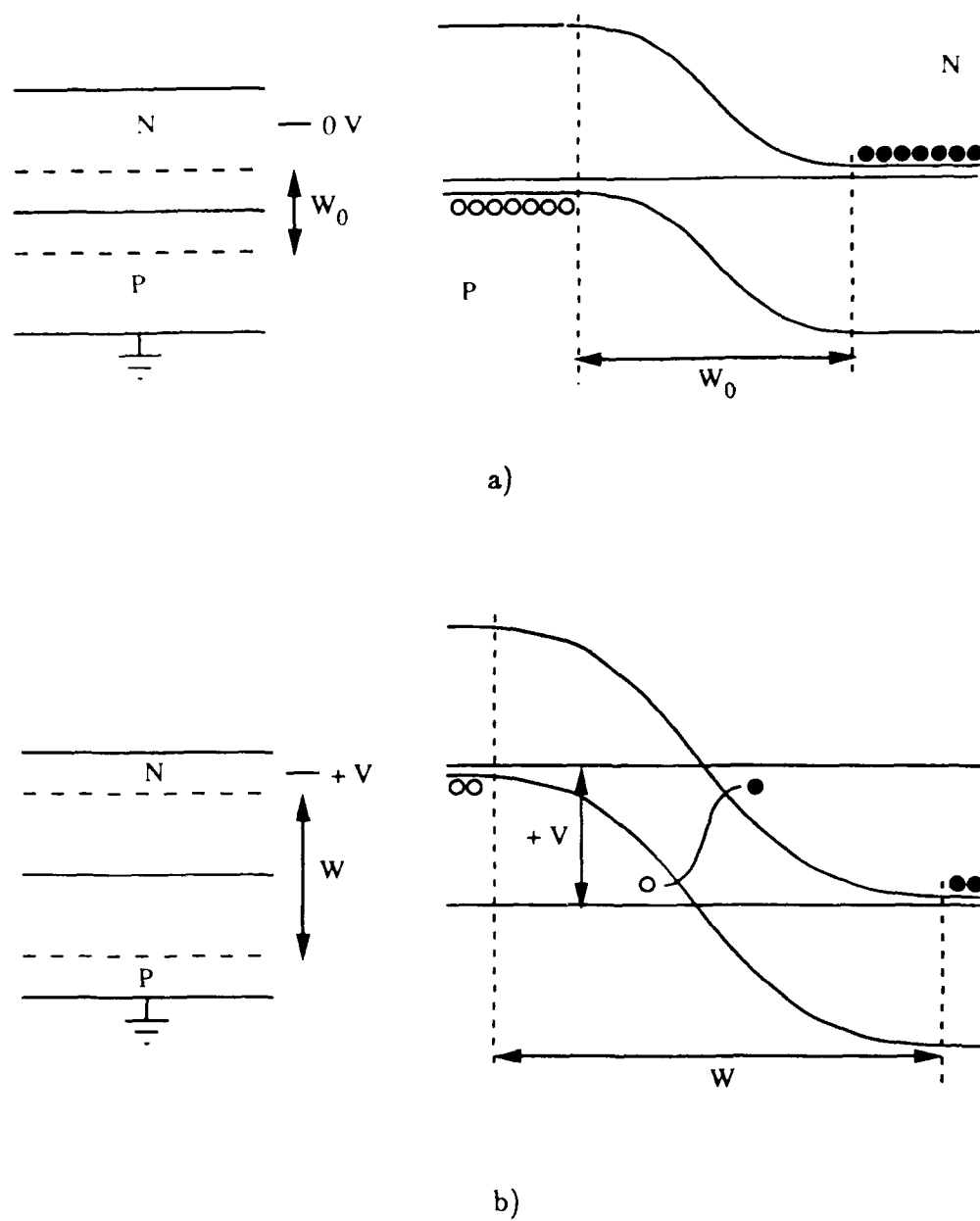


Figure 2.2 Logic "0" and logic "1" states of a PN-junction capacitor.

- a) Logic "0": equilibrium with N region at ground
- b) Logic "1": reverse bias with N region at a positive potential

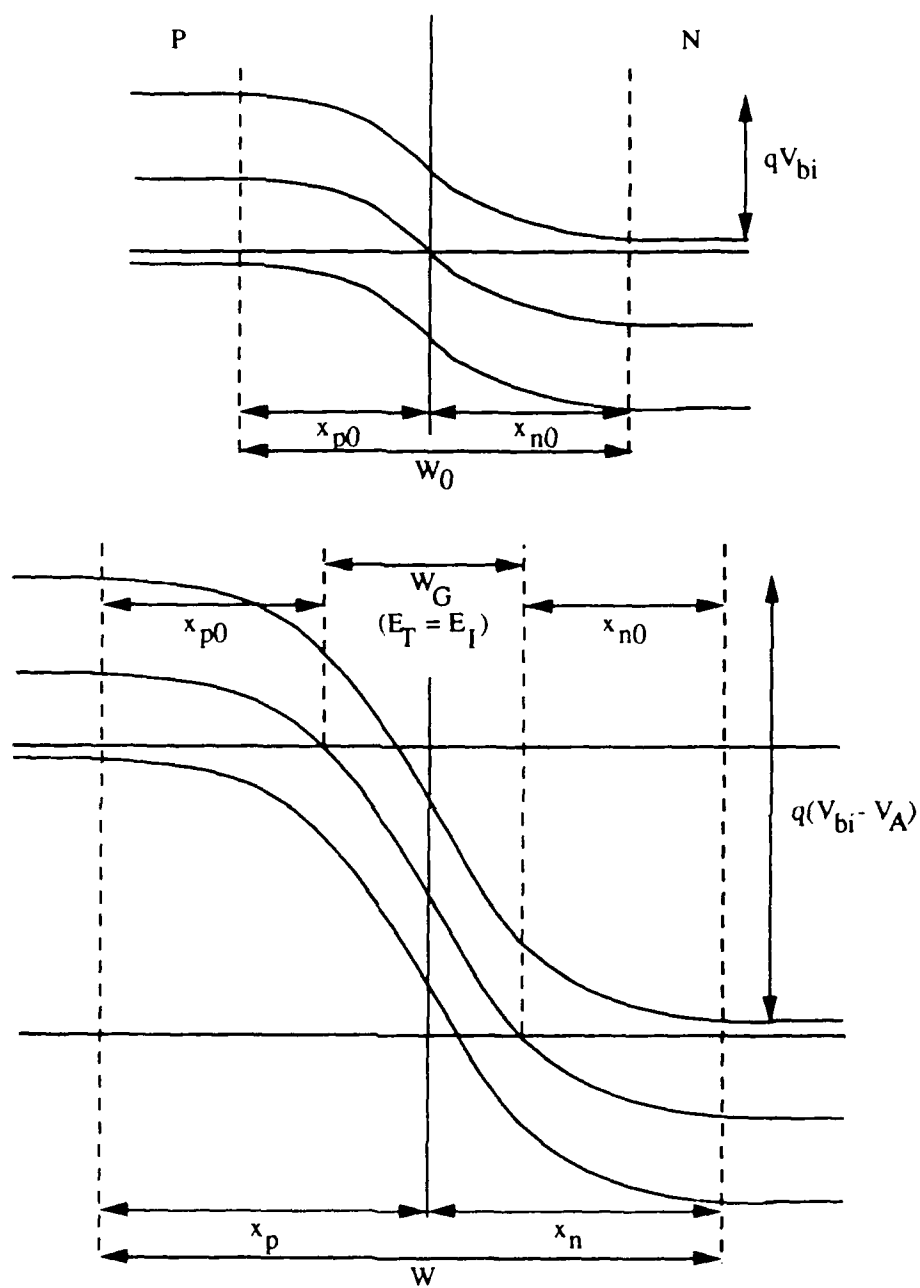


Figure 2.3 Definition of variables for the electrostatic solution of the PN junction.

$$V_{bi} = \frac{1}{q}(E_G + \eta_n + \eta_p) \quad [2.4]$$

where E_G is the bandgap of the material, and [28]:

$$\eta = \frac{\ln u}{1 - u^2} + \frac{(3\sqrt{\pi} u/4)^{2/3}}{1 + [0.24 + 1.08(3\sqrt{\pi} u/4)^{2/3}]^{-2}} \quad [2.5]$$

To calculate η_n , let $u = N_D/N_C$, and to calculate η_p , let $u = N_A/N_V$; where N_C and N_V are the effective densities of states in the conduction and valence bands.

The charge density per-unit-area stored on a reverse-biased GaAs PN junction is plotted in Figure 2.4 as a function of applied bias and doping level. The maximum charge density which can be stored on a junction is limited by breakdown due to avalanching or tunneling. Ideal planar diodes in GaAs require a charge density greater than $5 \text{ fC}/\mu\text{m}^2$ to cause breakdown, but this limit will not be reached for achievable doping densities using the small voltages employed in GaAs digital logic. A more conservative estimate for heavily-doped-diode charge storage density at GaAs voltage levels is $2\text{--}3 \text{ fC}/\mu\text{m}^2$. This is an order of magnitude lower than the theoretical maximum charge density achievable in silicon MOS capacitors [29], but it is comparable to charge densities actually obtained in planar Si cell designs [30].

2.2.2 Storage Time

A logic "1" is stored on a PN-junction capacitor by removing electrons from the floating N region, charging it to some positive potential. The storage time is determined by the rate at which the diode leakage currents resupply the missing electrons to the N region. Leakage currents in reverse-bias PN junctions are caused by generation of electron-hole pairs which separate into a free electron in the conduction band and a hole in the valence band. To contribute to the leakage current, the electron must make its way to the N side and the hole must get to the P side. Carriers generated within the depletion region move to their respective destinations by drifting in the electric field. Minority-carrier electrons generated in the neutral P region first diffuse to the depletion-region edge before being swept to the N side by the electric field. Similarly, minority-carrier holes generated in the neutral N region first diffuse and then drift to the P side. In this discussion, the current due to carriers generated within the depletion region will be referred to as depletion current. The current due to carriers generated outside the depletion edges will be referred to as diffusion current. Both the depletion current and the diffusion current will have components due to generation in the bulk of the device and at the perimeter.

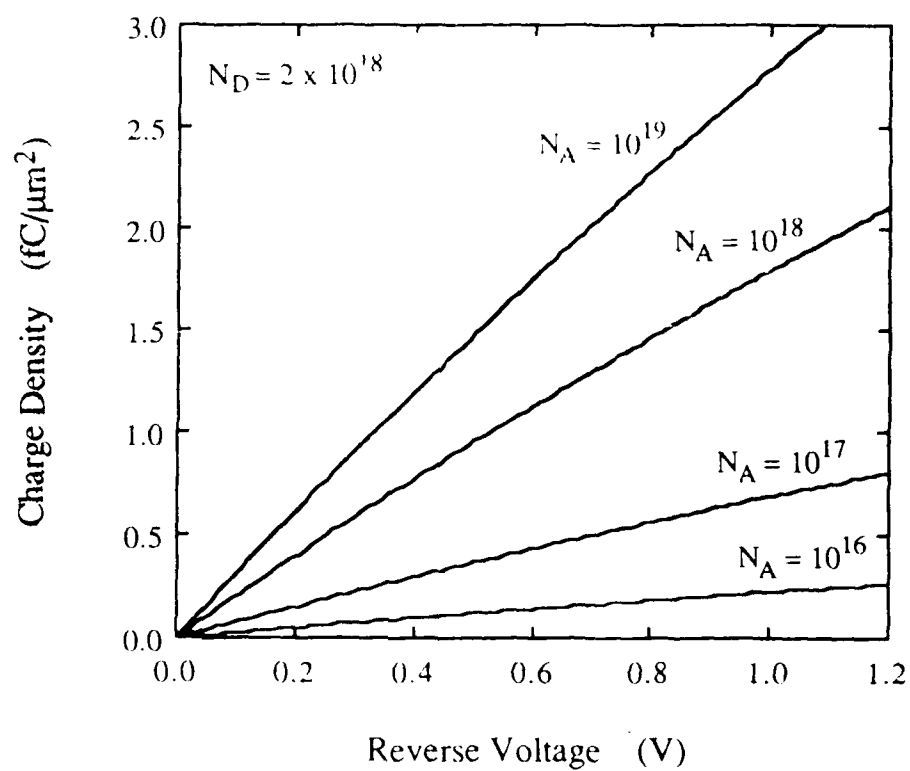


Figure 2.4 Charge density versus doping and bias in a PN-junction capacitor. In the calculation, $N_D = 2 \times 10^{18} \text{ cm}^{-3}$ and N_A assumes the values shown in the figure.

The following subsections develop expressions for the four components of the leakage current. Generation of electron-hole pairs under reverse-bias conditions occurs due to three mechanisms: thermal generation, generation due to radiation, and generation due to impact ionization. It is assumed that dynamic memory cells using PN-junction storage capacitors will be designed to employ electric fields small enough to prevent impact ionization. Then, in the absence of light or radiation from other sources, the generation rate will be dominated by thermal generation through recombination-generation centers. Shockley-Read-Hall theory [31] provides an expression for the generation rate in terms of carrier concentrations and R-G center parameters. It is necessary to know the positions of the band edges with respect to the quasi-Fermi levels in order to determine the carrier concentrations as a function of position. The familiar solution for the band edge positions in the bulk was discussed in Section 2.2.1 and illustrated in Figure 2.3. Section 2.2.2.3 will present a method for approximating the positions of the band edges along the surfaces which form the perimeter of the junction.

2.2.2.1 Bulk Depletion Current

Bulk generation is typically characterized by generation through distinct trap levels that are separated spatially and in energy. Under these conditions, the generation rate due to R-G centers at a single energy level is given by:

$$G_B = \frac{n_i^2 - np}{\tau_p(n + n_1) + \tau_n(p + p_1)} \quad \text{chp/cm}^3\text{s} \quad [2.6]$$

where n and p are the electron and hole concentrations, τ_n and τ_p are the generation lifetimes for electrons and holes, and n_1 and p_1 are parameters which depend on the energy level of the R-G center:

$$n_1 = n_i e^{(E_T - E_i)/kT} \quad [2.7]$$

$$p_1 = n_i e^{(E_i - E_T)/kT} \quad [2.8]$$

where E_T is the energy level of the traps. The carrier concentrations are specified by the positions of the quasi-Fermi levels, F_N and F_P :

$$n = n_i e^{(F_N - E_i)/kT} \quad [2.9]$$

$$p = n_i e^{(E_i - F_P)/kT} \quad [2.10]$$

$$np = n_i^2 e^{(F_N - F_P)/kT} = n_i^2 e^{-qV_{bi}/kT} \quad [2.11]$$

Equation 2.11 assumes that the quasi-Fermi levels remain constant across the

depletion region so that V_R is the reverse bias applied to the junction. Substituting equations 2.7 through 2.11 into equation 2.6 yields:

$$G_B = \frac{(1 - e^{-qV_R/kT})n_i}{\tau_p(e^{(E_T - E_i)/kT} + e^{(F_P - E_i)/kT}) + \tau_n(e^{(E_i - E_T)/kT} + e^{(E_i - F_N)/kT})} \quad [2.12]$$

For any appreciable reverse bias, $e^{-qV_R/kT} \ll 1$, and may be neglected. The denominator rapidly becomes large and reduces the generation rate for $F_N > E_T$ or $F_P < E_T$. The generation rate is taken to be zero outside the generation width, W_{GB} , defined as the region within which $F_P > E_T > F_N$. Inside W_{GB} , the exponential terms involving the quasi-Fermi levels are negligible, and:

$$G_B = \frac{n_i}{\tau_G} \quad \text{ehp/cm}^3\text{s} \quad [2.13]$$

where:

$$\tau_G \equiv \tau_p e^{(E_T - E_i)/kT} + \tau_n e^{(E_i - E_T)/kT} \quad [2.14]$$

Note that for generation due to traps located close to mid-gap, the generation width is the region in which E_i lies between F_N and F_P as shown in Figure 2.3, and:

$$\tau_G \approx \tau_p + \tau_n \quad [2.15]$$

Figure 2.5 illustrates the generation-width approximation. The solid curve indicates the generation rate calculated from equation 2.12, while the dashed rectangle represents the use of equation 2.13. Using the generation-width approximation, the bulk depletion current becomes:

$$I_{BDepl} = qAW_{GB} \frac{n_i}{\tau_G} \quad [2.16]$$

2.2.2.2 Bulk Diffusion Current

The bulk diffusion current is described by the ideal diode equation [32]:

$$I_{BDiff} = qA \left[\frac{D_N}{L_N} n_{p0} + \frac{D_P}{L_P} p_{n0} \right] (1 - e^{-qV_R/kT}) \quad [2.17]$$

where D_N , L_N , D_P , and L_P are the diffusion coefficients and diffusion lengths for electrons and holes. A is the junction area, and n_{p0} and p_{n0} are the equilibrium minority-carrier densities for electrons in P-type material and for holes in the N-type material. Equation 2.17 is derived by assuming low-level injection conditions and solving the minority-carrier diffusion equation in the neutral regions. For example, on the P side,

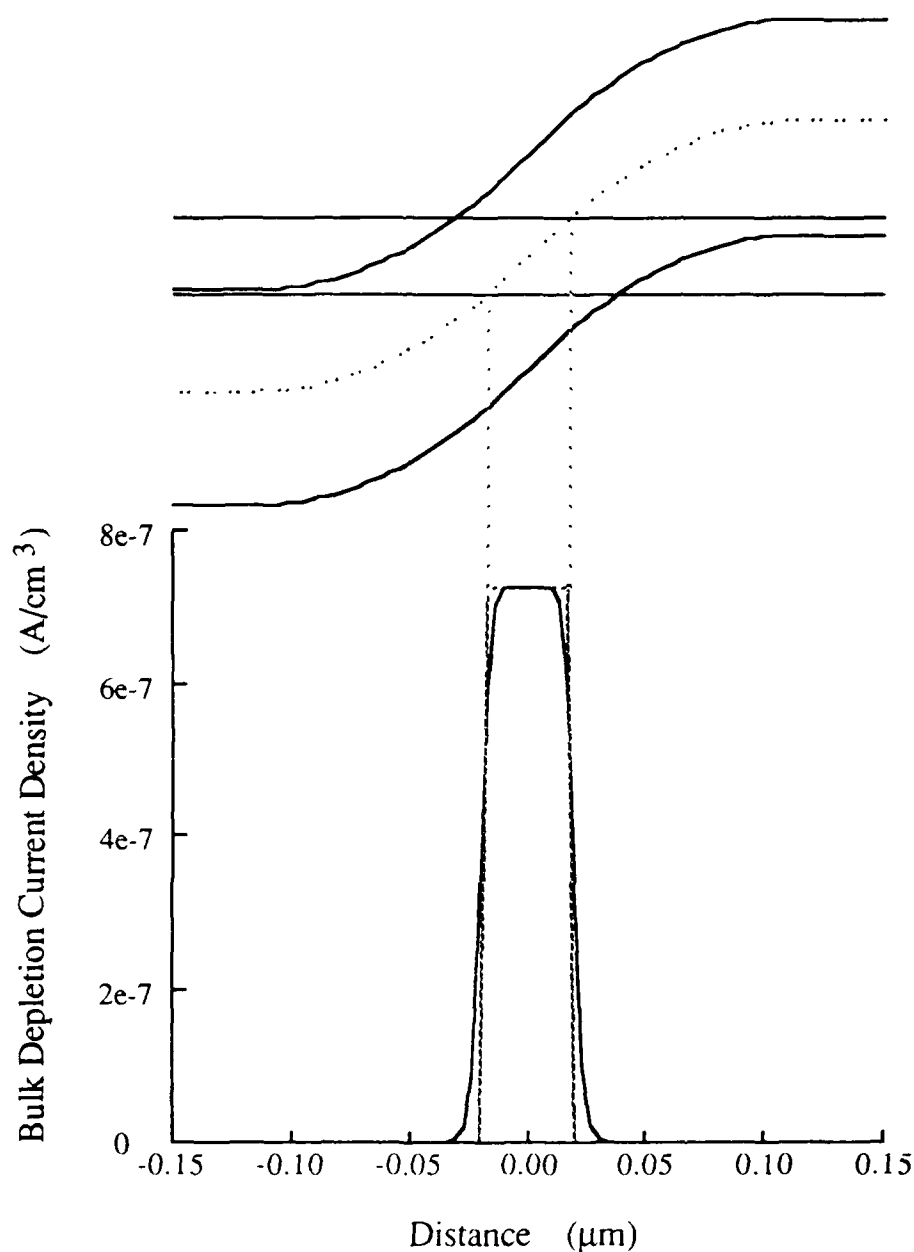


Figure 2.5 Calculated bulk depletion current density. The solid curve represents the full calculation from equation 2.12. The dashed rectangle represents the generation-width approximation of equation 2.13. Parameters: $N_A = N_D = 10^{17} \text{ cm}^{-3}$, $V_R = 0.5 \text{ V}$, $\tau_n = \tau_p = 10^{-7} \text{ s}$.

$$n_p(x) = n_{p0} + \Delta n_p \quad [2.18]$$

$$p_p(x) = p_{p0} + \Delta n_p \cong p_{p0} \quad [2.19]$$

$$D_p \frac{d^2 \Delta n_p}{dx^2} = -G_B \quad [2.20]$$

Substituting expressions 2.18 and 2.19 into 2.6 and ignoring negligible terms shows that the generation rate in the neutral region should be proportional to the change in the minority-carrier density:

$$G_B \cong \frac{-\Delta n_p}{\tau_n} \quad [2.21]$$

Substituting equation 2.21 into 2.20 and solving yields:

$$\Delta n_p(x) = n_{p0}(e^{-qV_R/kT} - 1)e^{-x/L_n} \quad [2.22]$$

where the diffusion length, L_n , is defined by:

$$L_n \equiv \sqrt{D_n \tau_n} \quad [2.23]$$

Equation 2.22 assumes the usual "long-base diode" boundary conditions:

$$\Delta n_p(0) = n_{p0}(e^{-qV_R/kT} - 1) \quad [2.24]$$

$$\Delta n_p(\infty) = 0 \quad [2.25]$$

where $x = 0$ is chosen to be the depletion region edge on the P side, with x increasing away from the junction. Finally, the current is entirely due to diffusion so:

$$J_n = qD_n \frac{d\Delta n_p}{dx} \quad [2.26]$$

All the electrons reaching the depletion region are assumed to make it across to the N side, so equation 2.26 is evaluated at $x = 0$:

$$J_n = q \frac{D_n}{L_n} n_{p0} (1 - e^{-qV_R/kT}) \quad [2.27]$$

Equation 2.27 and the complementary equation due to holes diffusing to the depletion edge on the N side combine to yield equation 2.17.

In GaAs, the bulk diffusion current is negligibly small because the minority-carrier densities are so low:

$$n_{p0} = \frac{n_i^2}{p_{p0}} = \frac{n_i^2}{N_A} \quad [2.28]$$

$$p_{n0} = \frac{n_i^2}{n_{n0}} = \frac{n_i^2}{N_D} \quad [2.29]$$

At room temperature, $n_i^2 \cong 3 \times 10^{12} \text{ cm}^{-6}$, while the lowest achievable doping densities are approximately 10^{14} cm^{-3} . Because there are essentially zero minority carriers in equilibrium, reverse biasing the junction can produce no minority-carrier concentration gradient in the neutral bulk, and the bulk diffusion current may be ignored.

2.2.2.3 Perimeter Depletion Current

The perimeter of a semiconductor device is characterized by a distribution of traps which is essentially continuous in energy, so that computing the generation rate requires integration of contributions from all energies from valence band to conduction band:

$$G_P = \int_{E_v}^{E_c} \frac{n_i^2 - n_s p_s}{(n_s + n_{1s})/c_{ps} + (p_s + p_{1s})/c_{ns}} D_{IT}(E) dE \quad (\text{ehp/cm}^2\text{s}) \quad [2.30]$$

where n_s and p_s are the carrier concentrations evaluated at the surface that defines the junction's perimeter. $D_{IT}(E)$ is the density of R-G centers at the surface in $(\text{cm}^2\text{eV})^{-1}$, and c_{ns} and c_{ps} are the capture coefficients for electron and hole traps at the surface in cm^3/s . The surface carrier concentrations are determined by equations analogous to 2.9 and 2.10:

$$n_s = n_i e^{(F_{Ns} - E_i)/kT} \quad [2.31]$$

$$p_s = n_i e^{(E_i - F_{Ps})/kT} \quad [2.32]$$

where F_{Ps} and F_{Ns} represent the quasi-Fermi levels at the surface.

Using equations 2.31 and 2.32 requires solving for the positions of the band edges with respect to the quasi-Fermi levels at the surface. Rigorously, this means solving the two-dimensional Poisson equation consistently with a variable interface charge. The problem can be treated approximately by solving one-dimensional Poisson equations normal to the surface [33].

For example, first consider the simpler case of calculating the positions of the band edges with respect to the Fermi level at the surface of a uniformly-doped N-type region in equilibrium. The density of electrons occupying interface states is given by:

$$n_T = q \int_{E_{v_s}}^{E_{c_b}} f_T(E, E_F) D_{IT}(E) dE \quad (\text{cm}^{-2}) \quad [2.33]$$

where $f_T(E, E_F)$ is the probability of occupancy for a surface state with energy E . To maintain charge neutrality, the surface charge must balance the charge in the bulk, which is assumed to be depleted to a depth y_N :

$$qN_D y_N = q(n_T - n_{TN}) \quad [2.34]$$

where n_{TN} is the trapped-electron density required to make the surface neutral. The left term of equation 2.34 represents the bulk depletion charge, balanced by the surface charge on the right. As indicated in Figure 2.6, Poisson's equation requires:

$$\frac{qN_D}{2\epsilon} y_N^2 = \frac{E_{C_s} - E_{C_b}}{q} \quad [2.35]$$

where E_{C_b} is the position of the conduction band edge in the neutral bulk. E_{C_b} is fixed with respect to the Fermi level by the doping through equation 2.9. The position of E_{C_s} with respect to the Fermi level controls $f_T(E, E_F)$. For specified $D_{IT}(E)$, n_{TN} , and $f_T(E, E_F)$, there is a unique solution to equations 2.34 and 2.35 for E_{C_s} .

To apply this technique to solve for the band edge positions where a PN junction intersects the surface, the following additional assumptions are imposed:

- 1) The surface generation velocity is small enough so that the majority-carrier quasi-Fermi levels do not bend significantly normal to the surface in regions in which the majority-carrier concentration at the surface exceeds the intrinsic concentration.
- 2) The effect of electric fields parallel to the surface on the one-dimensional Poisson solutions normal to the surface is ignored. This is equivalent to assuming that the electric field due to ionized dopants within the distance y_N from the surface is entirely terminated by surface charge.
- 3) The neutral energy level, E_N , defined by:

$$n_{TN} = \int_{E_v}^{E_N} D_{IT}(E) dE \quad [2.36]$$

is not far from mid-gap.

Assumption 1 means that the positions of the majority-carrier quasi-Fermi levels at the surface are fixed by the applied reverse bias, just as in the bulk.

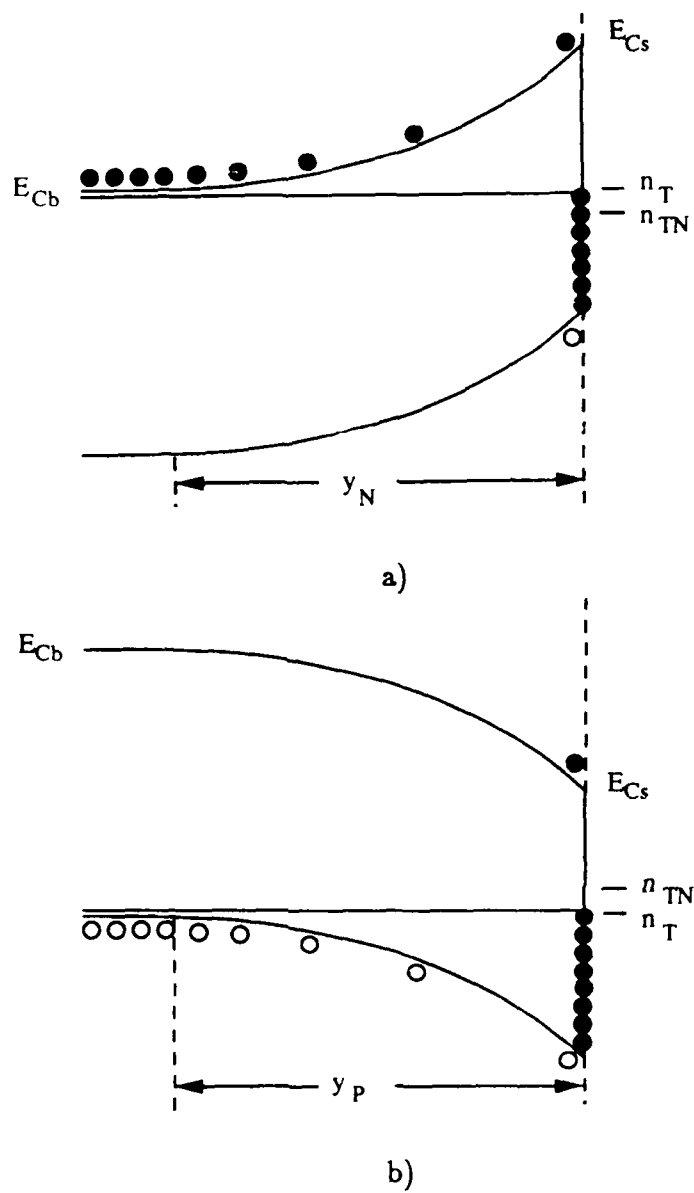


Figure 2.6 Band diagrams normal to pinned surfaces in equilibrium.

a) N-type semiconductor

b) P-type semiconductor

This assumption can be checked after the solution is complete by calculating the surface generation current density, which sets the maximum bending of the quasi-Fermi levels normal to the surface. The second assumption technically violates the two-dimensional Poisson equation, but under the condition specified by assumption 3, the band bending normal to the surface will be shown to be small in the critical region close to the junction.

The probability of occupancy function for surface states is derived from Shockley-Read-Hall theory by requiring the steady-state electron and hole generation rate to balance:

$$C_n = e_n n_T - c_n n_s p_T = (e_n f_T - c_n n_s (1 - f_T)) D_{IT}(E) \quad [2.37]$$

$$G_p = e_p p_T - c_p p_s n_T = (e_p (1 - f_T) - c_p p_s f_T) D_{IT}(E) \quad [2.38]$$

where e_n , e_p , c_n , and c_p are the electron and hole capture and emission coefficients at the surface, and n_T and p_T are the numbers of filled and vacant surface states with energy E . In equations 2.37 and 2.38, the "s" subscripts and the dependence of f_T on E , F_{Ns} , and F_{Ps} have been suppressed. Equating 2.37 and 2.38 yields:

$$f_T = \frac{c_n n_s + e_p}{c_n n_s + e_n + c_p p_s + e_p} \quad [2.39]$$

The principle of the detailed balance equation requires that:

$$e_n = n_1 c_n \quad [2.40]$$

$$e_p = p_1 c_p \quad [2.41]$$

where n_1 and p_1 are defined as in equations 2.7 and 2.8. Using 2.40 and 2.41,

$$f_T = \frac{c_n n_s + c_p p_1}{c_n (n_s + n_1) + c_p (p_s + p_1)} \quad [2.42]$$

Consider a region where E_i lies between the split quasi Fermi levels in reverse bias. Within such a region, equation 2.42 shows that for surface states of energy E greater than both F_{Ns} and F_{Ps} , f_T goes to zero because n_1 is large. For states with energy less than both F_{Ns} and F_{Ps} , f_T goes to one because p_1 dominates:

$$f_T = 0, \quad E > F_{Ps} \geq F_{Ns} \quad [2.43]$$

$$f_T = 1, \quad F_{Ps} \geq F_{Ns} > E \quad [2.44]$$

For states with energies between F_{Ns} and F_{Ps} , equation 2.43 becomes:

$$f_T = \frac{c_p P_1}{c_n n_1 + c_p P_1} \quad F_{Ps} > E > F_{Ns} \quad [2.45]$$

For a density of interface states that is uniform in energy, equation 2.33 can be written as:

$$n_T = q \frac{N_{IT}}{E_G} \int_{E_{Vs}}^{E_{Cs}} f_T(E, F_{Ns}, F_{Ps}) dE \quad [2.46]$$

where N_{IT} is the total number of interface states per unit area. Using approximations 2.43 through 2.45, equation 2.46 becomes:

$$n_T = q N_{IT} \frac{F_{Ns} - E_{Vs}}{E_G} + q \frac{N_{IT}}{E_G} \int_{F_{Ns}}^{F_{Ps}} \frac{c_p P_1}{c_n n_1 + c_p P_1} dE \quad [2.47]$$

To evaluate the integral in 2.47, substitute 2.9 and 2.10 and note that:

$$\int_{F_{Ns}}^{F_{Ps}} \frac{c_p P_1}{c_n n_1 + c_p P_1} dE = \int_{F_{Ns}}^{F_{Ps}} \frac{c_p e^{2(E_i - E)/kT}}{c_n + c_p e^{2(E_i - E)/kT}} dE \quad [2.48]$$

Integrating,

$$\int_{F_{Ns}}^{F_{Ps}} f_T dE = -\frac{kT}{2} \ln \left(\frac{c_n + c_p e^{2(E_i - F_{Ps})/kT}}{c_n + c_p e^{2(E_i - F_{Ns})/kT}} \right) \quad [2.49]$$

Assuming that c_n is approximately equal to c_p , the solution to equation 2.48 for a region where E_i lies between F_{Ns} and F_{Ps} becomes:

$$\int_{F_{Ns}}^{F_{Ps}} f(T) dE = E_i - F_{Ns} \quad , \quad F_{Ps} > E_i > F_{Ns} \quad [2.50]$$

Combining equations 2.47 and 2.50 yields:

$$n_T = q \frac{N_{IT}}{E_G} (E_i - E_{Vs}) \quad , \quad F_{Ps} > E_i > F_{Ns} \quad [2.51]$$

Here F_{Ns} is assumed to exceed E_{Vs} , and F_{Ps} is assumed to be less than E_{Cs} . If $F_{Ns} < E_{Vs}$, F_{Ns} should be replaced by E_{Vs} in equations 2.43 through 2.51. Similarly, if $F_{Ps} > E_{Cs}$, F_{Ps} should be replaced by E_{Cs} .

The majority-carrier quasi-Fermi levels must be pinned above and below E_i in the neutral N and P regions, respectively. If assumption 1 holds, then there must exist a region at the surface for which $F_{Ps} > E_i > F_{Ns}$. Note that, using the approximations of uniform interface-state distribution, equation 2.36

becomes:

$$n_{TN} = \frac{qN_{IT}}{E_G}(E_N - E_{Vs}) \quad [2.52]$$

If assumption 3 also holds, so that $E_N \cong E_i$, then within the region where E_i is between the quasi-Fermi levels, equations 2.51 and 2.52 give:

$$n_T \cong n_{TN} \quad , \quad F_{Ps} > E_i > F_{Ns} \quad [2.53]$$

Equations 2.34 and 2.35 then require that the bands be flat normal to the surface in this region. This relieves the violation of the 2-D Poisson equation implied by assumption 2, by requiring the bands in this region to bend along the surface in the same way as they do in the bulk.

The region within which $F_{Ps} > E_i > F_{Ns}$ is defined as the perimeter generation width, W_{GP} . Within W_{GP} , n_s and p_s are small, and equation 2.30 can be written as:

$$G_P = n_i \int_{E_v}^{E_c} \frac{c_{ns}c_{ps}}{c_{ns}e^{(E - E_i)/kT} + c_{ps}e^{(E_i - E)/kT}} D_{IT} dE \quad [2.54]$$

$$G_P = s_0 n_i \quad [2.55]$$

where the parameter s_0 defined by equations 2.54 and 2.55 is the surface generation velocity in cm/s. Outside W_{GP} , the generation rate drops off as n_s or p_s increases. Because the surface equilibrium carrier concentrations are lower than in the bulk, the generation rate outside W_{GP} decreases more slowly. Neglecting the generation outside W_{GP} , the perimeter depletion current can be expressed as:

$$I_{PDepl} = qs_0 n_i W_{GP} P \quad [2.56]$$

where P is the length of the perimeter of the device.

Numerically, it is possible to solve equations 2.34 and 2.35 using equation 2.42 directly rather than approximating the probability of occupancy function. A typical solution for the bands at the surface is shown in Figure 2.7. Note the high electric field, corresponding to the band bending in the bulk, within W_{GP} . Also shown in the figure is the perimeter generation rate calculated directly from equation 2.30, and approximated by equation 2.55. Comparing Figure 2.5 with Figure 2.7, it is apparent that the generation-width approximation is less accurate in describing the perimeter than it is for the bulk.

It is implied by both the assumptions and the solution that the majority of the current flow is parallel to the surface from the N region to the P region.

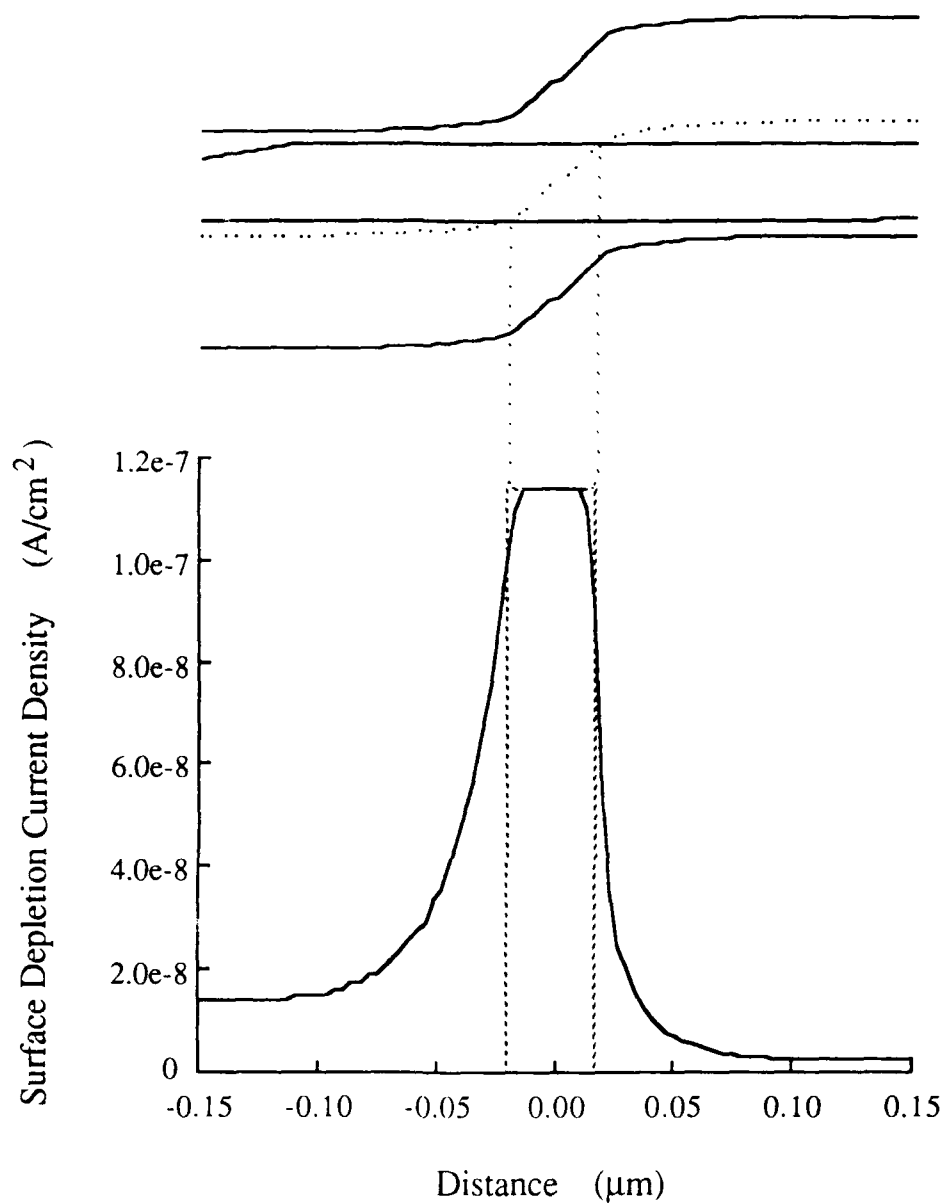


Figure 2.7 Calculated perimeter depletion current density. The solid curve represents the full calculation from equation 2.30. The dashed rectangle represents the generation-width approximation of equation 2.55. Parameters: $N_A = N_D = 10^{17} \text{ cm}^{-3}$, $V_R = 0.5 \text{ V}$, $N_{IT} = 10^{13} \text{ cm}^{-2}$, $s_0 = 4 \cdot 10^5 \text{ cm/s}$.

The drift and diffusion components normal to the surface approximately cancel in steady state, just as they do in equilibrium. This helps to justify assumption 1, which requires the majority-carrier quasi-Fermi levels to be flat from the bulk to the surface. An upper limit on the maximum majority-carrier quasi-Fermi level bending from the bulk to the surface can be calculated by assuming that all the generation current flows directly into the bulk. The current density is related to the quasi-Fermi level bending by:

$$J_{N,y} = q\mu_n n \frac{dF_N}{dy} \quad [2.57]$$

where y is the direction normal to the surface and μ_n is the electron mobility. Integrating both sides gives:

$$\int_0^y dF_N = F_N(y) - F_N(0) = \int_0^y \frac{J_{N,y}}{q\mu_n n} dy \quad [2.58]$$

To calculate an upper limit for the bending of the majority-carrier quasi-Fermi level, choose maximum values for y and $J_{N,y}$ and set n to its minimum. On the N side, outside of W_{GP} , the minimum electron concentration is the intrinsic concentration. The maximum current density is given by equation 2.55. The maximum value for y can be taken as the equilibrium surface depletion width in the neutral region. Using these values, equation 2.58 gives:

$$\left| F_{N\text{bulk}} - F_{Ns} \right|_{\text{max}} = \int_0^{y_n} \frac{q s_0 n_i}{q\mu_n n_i} dy = \frac{s_0}{\mu_n} y_{N0} \quad [2.59]$$

Similarly,

$$\left| F_{P\text{bulk}} - F_{Ps} \right|_{\text{max}} = \frac{s_0}{\mu_p} y_{P0} \quad [2.60]$$

Using the typical parameter values of Figure 2.7, these upper limits are less than one kT .

2.2.2.4 Perimeter Diffusion Current

As discussed in Section 2.2.2.2, in the bulk of a GaAs diode the diffusion current is negligibly small because the minority-carrier concentrations are so low. At the perimeter of the device, the pinning of the majority-carrier quasi-Fermi levels near mid-gap results in increased minority-carrier densities and increased diffusion current. Figure 2.8 compares the expected carrier concentrations at the surface and in the bulk of a GaAs diode. The lowest diagram in the figure indicates that, when the diode is reverse biased, a minority-carrier gradient will exist along the surface outside the region where

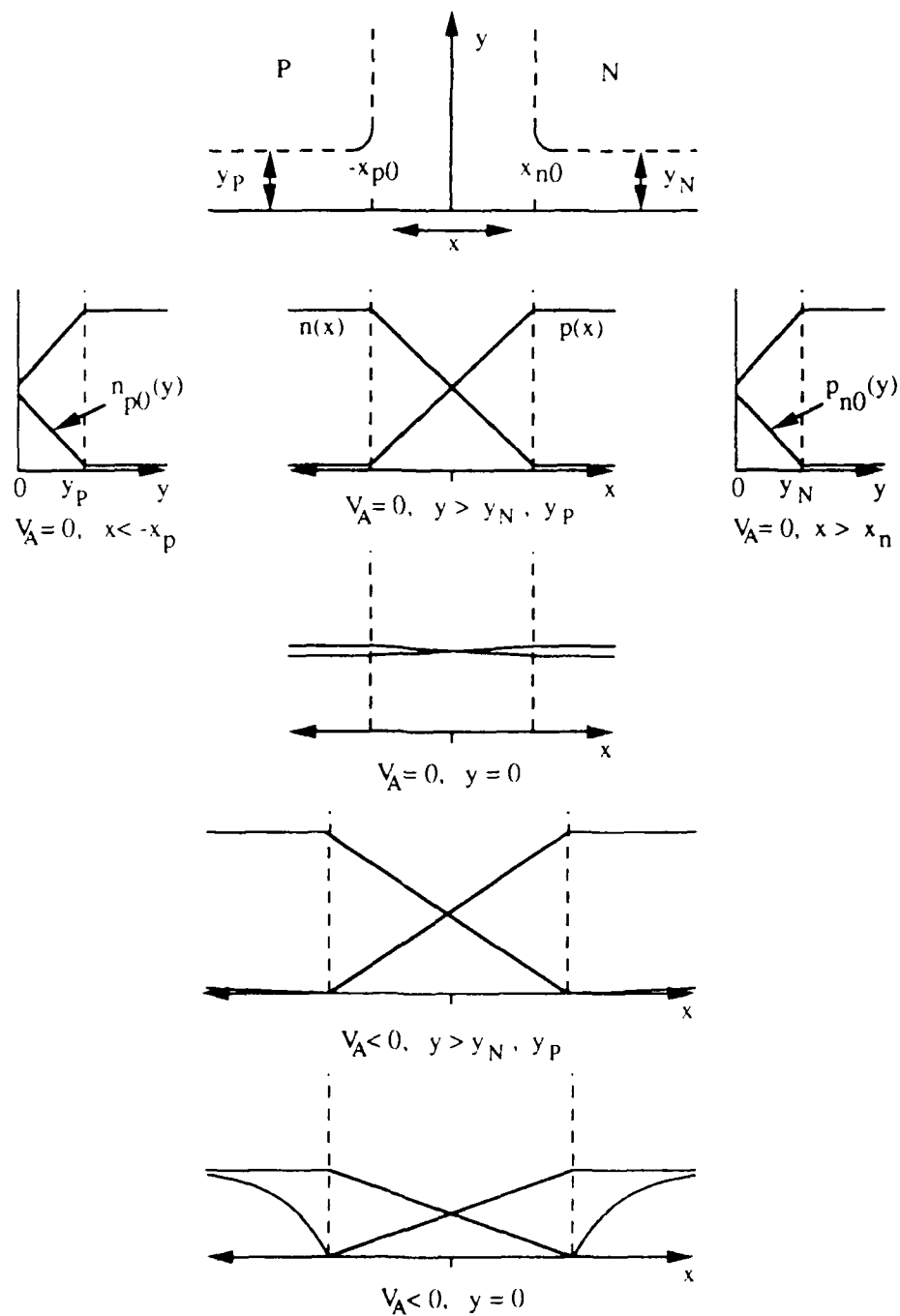


Figure 2.8 Carrier concentrations versus distance at the surface and in the bulk.

there is an electric field parallel to the surface. This gradient in the minority-carrier concentration results in a surface diffusion current.

The surface diffusion current can be derived approximately by solving the minority-carrier diffusion equation parallel to the surface in the regions where the electric field component along the surface goes to zero. Technically, an electric field component normal to the surface exists everywhere, which violates the assumptions of the minority-carrier diffusion equation. As described in the previous subsection, it can be assumed that the drift and diffusion currents normal to the surface cancel each other in steady state, just as they do in equilibrium.

As is indicated in Figure 2.6, the minority carriers are confined near the surface in approximately triangular potential wells. In order to express the perimeter diffusion current in terms of the surface generation rate, the current density along the surface is assumed constant from the surface out to some width W_n or W_p , on the P and N sides, respectively [33]. Then, choosing x along the surface normal to the junction, the minority diffusion equation for electrons becomes:

$$D_N \frac{d^2 n_{ps}}{dx^2} = \frac{G_p}{W_n} \quad [2.61]$$

where n_{ps} is the minority-carrier electron concentration evaluated at the surface, and G_p is the perimeter generation rate in the regions where the electric field along the surface is zero and where the quasi-Fermi levels are still split. For example, consider the region $y = 0$, $x < -x_p$ in the lowest diagram of Figure 2.8. In this region,

$$p_s = p_{ps0} \quad [2.62]$$

$$n_s = n_{ps0} - \Delta n_{ps}(x) \quad [2.63]$$

where p_{ps0} and n_{ps0} are the hole and electron concentrations at the surface at equilibrium. For typical equilibrium surface carrier concentrations and typical trap parameters, using equations 2.62 and 2.63 in the general surface-generation-rate expression (equation 2.30), produces a generation rate which is independent of Δn_{ps} for $n_{ps} \ll n_i$, and which is proportional to Δn_{ps} for n_{ps} comparable to n_i . Equation 2.61 requires different forms for the x -dependence of Δn_{ps} when G_p is a constant and when G_p is proportional to Δn_{ps} . Surface generation velocities can be defined for each case:

$$G_p = s'_n n_i = \left[\int_{E_v}^{E_c} \frac{n_i D_{IT} dE}{n_{1s}/c_{ps} + (p_{ps0} + p_{1s})/c_{ns}} \right] n_i, \quad n_{ps} \ll n_i \quad [2.64]$$

$$G_p = s_n \Delta n_{ps} = \left[\int_{E_v}^{E_c} \frac{p_{ps0} D_{IT} dE}{n_{1s}/c_{ps} + (p_{ps0} + p_{1s})/c_{ns}} \right] \Delta n_{ps}, \quad n_s \lesssim n_i \quad [2.65]$$

Approximating the generation rate as a constant for all Δn_{ps} and substituting 2.64 into 2.61 produces a parabolic solution for $\Delta n_{ps}(x)$:

$$D_n \frac{d^2 \Delta n_{ps}}{dx^2} = s'_n n_i \quad [2.66]$$

$$\Delta n_{ps}(x) = -C_n (x - L'_{ns})^2 \quad [2.67]$$

where evaluating boundary conditions gives:

$$C_n = \frac{s'_n n_i}{2D_n W_n} \quad [2.68]$$

$$L'_{ns} = \left(\frac{2D_n W_n}{s'_n n_i} n_{ps0} (1 - e^{-qV_R/kT}) \right)^{1/2} \quad [2.69]$$

As usual, $x = 0$ has been chosen to be at the edge of the surface electric field region, with x increasing away from the junction.

Treating the generation rate as proportional to Δn_{ps} for all Δn_{ps} produces the familiar exponential form:

$$D_n \frac{d^2 \Delta n_{ps}}{dx^2} = s_n \Delta n_{ps} \quad [2.70]$$

$$\Delta n_{ps}(x) = n_{ps0} (e^{-qV_R/kT} - 1) e^{-x/L_{ns}} \quad [2.71]$$

$$L_{ns} = (D_n \frac{W_n}{s_n})^{1/2} \quad [2.72]$$

The minority-carrier distributions calculated from equations 2.67 and 2.71 are shown in Figure 2.9. The calculation uses $N_A = N_D = 10^{17} \text{ cm}^{-3}$, $s_0 = 4 \times 10^5 \text{ cm/s}$, $N_{IT} = 10^{13} \text{ cm}^{-2}$, and 0.5 V reverse bias. In the real solution, the minority-carrier concentration will initially vary rapidly with distance along the path of the parabolic solution. When the difference in the quasi-Fermi levels becomes small, the variation of the minority-carrier concentration will follow the exponential solution.

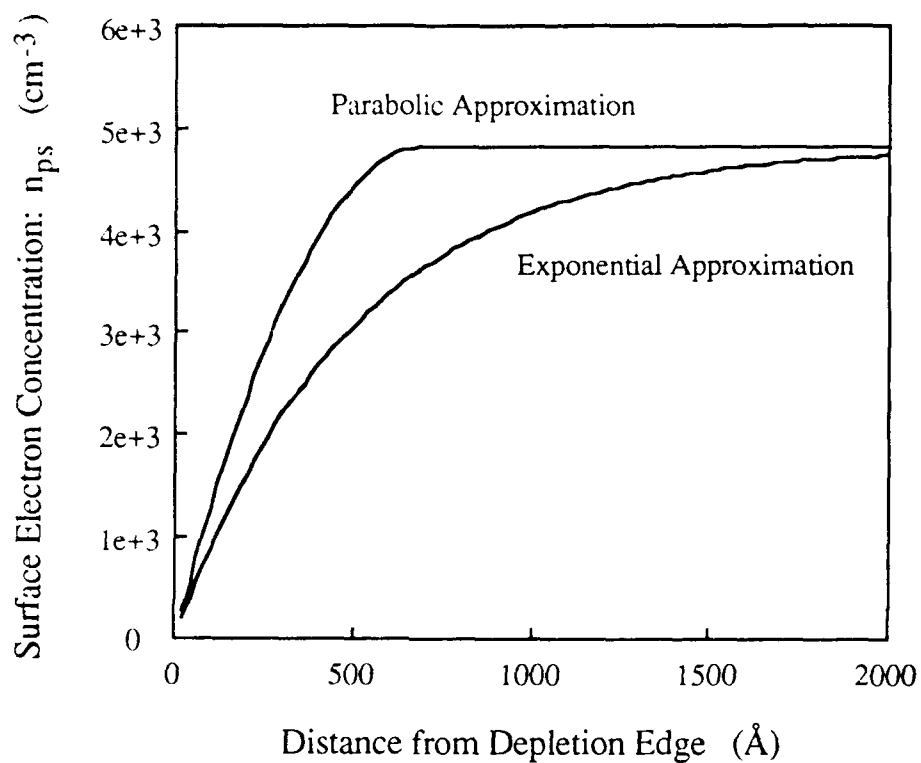


Figure 2.9 Surface minority-carrier electron concentration in the neutral region versus distance from the depletion edge. The two curves are calculated using the parabolic approximation of equation 2.67 and the exponential approximation of equation 2.71.

Despite the differences in the minority-carrier distributions described by equations 2.67 and 2.71, the corresponding diffusion currents are quite similar. The diffusion current is calculated from:

$$J_{ns} = qD_n \frac{dn_s}{dx} \quad [2.73]$$

Equation 2.73 is evaluated at the edge of the region where there is an electric field along the surface due to the junction. Using the parabolic solution of equation 2.67,

$$J'_{ns} = qD_n \left(\frac{2s'_n}{D_n w_n} n_i n_{ps0} (1 - e^{-qV_R/kT}) \right)^{1/2} \quad [2.74]$$

Using the exponential solution of equation 2.71,

$$J_{ns} = qD_n \left(\frac{s_n}{D_n w_n} \right) n_{ps0} (1 - e^{-qV_R/kT}) \quad [2.75]$$

The currents calculated from equations 2.74 and 2.75 are plotted versus reverse-bias voltage in Figure 2.10. The true solution for the perimeter diffusion current will lie somewhere between the two approximations. Also plotted in Figure 2.10 is the calculated perimeter depletion current using the same interface-state parameters. It is apparent that, using the specified parameter values, the perimeter depletion current will dominate the perimeter diffusion current in a GaAs homojunction except at very low biases. The surface recombination velocity used in the calculation is 10^5 cm/s, which is a relatively low value. For larger surface recombination velocities, the ratio of depletion current to diffusion current becomes even larger.

The fact that the parabolic and exponential minority-carrier approximations lead to different minority-carrier quasi-Fermi level positions but to nearly the same current points out that the exact positions of the minority-carrier quasi-Fermi levels are not critical to leakage calculations. The use of V_R in equations 2.74 and 2.75 implicitly assumes that the minority-carrier quasi-Fermi levels remain flat across the whole region in which the bands are bent along the surface. In the presence of strong surface generation, this assumption is unlikely to be true. The surface minority-carrier quasi-Fermi levels may instead begin to bend back toward the majority-carrier quasi-Fermi levels before E_x goes to zero. In this case, V_R in equations 2.74 and 2.75 would be replaced by the difference in the quasi-Fermi levels that remained at $x=0$. However, as long as the quasi-Fermi levels are still split by a few kT , the calculated diffusion current will remain essentially unchanged.

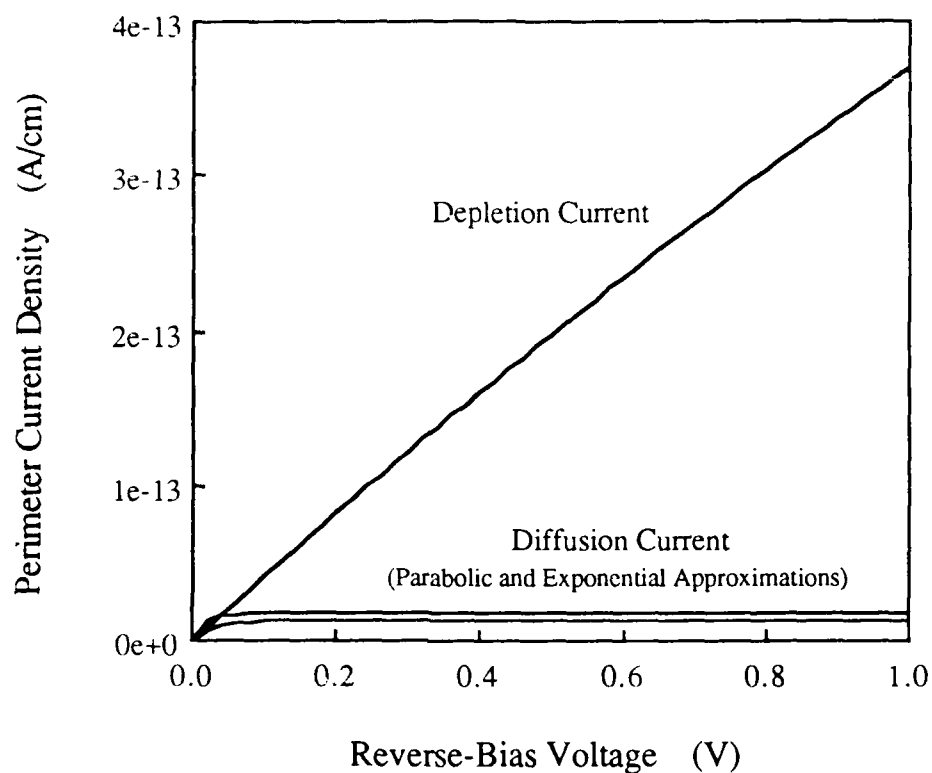


Figure 2.10 Calculated perimeter depletion and diffusion currents. The upper curve represents the perimeter depletion current calculated from equation 2.56. The lower two curves represent the perimeter diffusion current as approximated by equations 2.74 and 2.75. The parameters used in the calculation are $N_A = N_D = 10^{17} \text{ cm}^{-3}$, $N_{IT} = 10^{13} \text{ cm}^{-2}$, and $s_0 = 10^5 \text{ cm/s}$.

2.2.2.5 Bulk and Perimeter Limited Storage Time Expressions

To derive an expression for charge recovery as a function of time, start with equation 2.1, and define the total charge, Q_T , as the charge density times the area of the junction:

$$Q_T = qN_A(x_p - x_{p0})A \quad [2.76]$$

Charge neutrality requires that:

$$x_p - x_{p0} = \frac{N_D}{N_A + N_D}(W - W_0) \quad [2.77]$$

so,

$$Q_T = q \frac{N_A N_D}{N_A + N_D}(W - W_0)A \quad [2.78]$$

Neglecting diffusion currents, the charge recovery rate can be expressed as:

$$\frac{dQ_T}{dt} = -qG_B W_{GB}A - qG_P W_{GP}P \quad [2.79]$$

As is illustrated in Figure 2.3, in the bulk of a symmetrically doped diode, the generation width for states near mid-gap is the difference between the total depletion width and its equilibrium value. As described in Subsection 2.2.2.3, the generation width at the surface is approximately equal to the generation width in the bulk:

$$W_G = W_{GP} = W_{GB} = W - W_0 \quad (N_A = N_D) \quad [2.80]$$

Letting $N_A = N_D = N_B$ in equation 2.78 gives:

$$Q_T = \frac{1}{2}qAN_B W_G \quad [2.81]$$

Substituting 2.81 into 2.79:

$$\frac{dQ_T}{dt} = \frac{-2}{N_B}(G_B + G_P \frac{P}{A})Q_T \quad [2.82]$$

Solving,

$$Q_T = Q_{T0} e^{\frac{-2(G_B + G_P P/A)t}{N_B}} \equiv Q_{T0} e^{\frac{-t}{\tau_S}} \quad [2.83]$$

where τ_S , defined as the storage time constant, is given by:

$$\tau_S = \frac{N_B}{2(G_B + G_P P/A)} \quad [2.84]$$

Substituting expression 2.13 shows that for a bulk dominated device,

$$\tau_B = \frac{N_B \tau_G}{2n_i} \quad [2.85]$$

While expression 2.55 shows that for a perimeter dominated device:

$$\tau_P = \frac{N_B}{2s_0 n_i} \frac{A}{P} \quad [2.86]$$

Equations 2.85 and 2.86 show why reasonable storage-time performance may be expected from GaAs PN-junction capacitors even though GaAs has much shorter minority-carrier lifetimes and higher surface recombination velocity than Si. The intrinsic carrier concentration in the denominators of the storage-time expressions offsets the effects of small τ_0 and large s_0 . For a non-degenerate semiconductor:

$$n_i = \sqrt{N_C N_V} e^{-E_G/2kT} \quad [2.87]$$

where N_C and N_V are the density of states in the conduction and valence bands, and E_G is the energy bandgap of the material. The wider bandgap of GaAs compared to Si results in a much smaller intrinsic carrier concentration, which offsets GaAs's short τ_0 and large s_0 , resulting in a relatively long storage time.

2.2.3 Experimental Demonstration of PN-Junction Capacitors

The storage times achievable in GaAs PN-junction capacitors have been demonstrated using the structure shown in Figure 2.11 [34]. The film was grown by molecular beam epitaxy using a Perkin-Elmer Phi-400 MBE machine. Individual devices were isolated by wet chemical etching in 1:8:40 $H_2SO_4:H_2O_2:DI$. The top contact was made by thermally evaporating AuZn and patterning by lift-off. Contact to the back side was made through indium alloyed during MBE growth.

The structure consists essentially of an N-type potential well between two P-type regions, forming two back-to-back PN junctions. Figure 2.12 shows a simplified energy-band diagram for the device. As is indicated in the figure, applying a negative bias to the top contact will forward bias the upper PN junction, removing electrons from the floating N region. The N region charges to some positive potential, and when the bias is returned to zero, there is no source of electrons to refill the well except for the diode leakage currents.

The capacitance-voltage characteristics of the structure is shown in Figure 2.13. The approximate physical symmetry of the film results in the symmetry of the CV curve. As the applied bias is swept slowly in either direction from

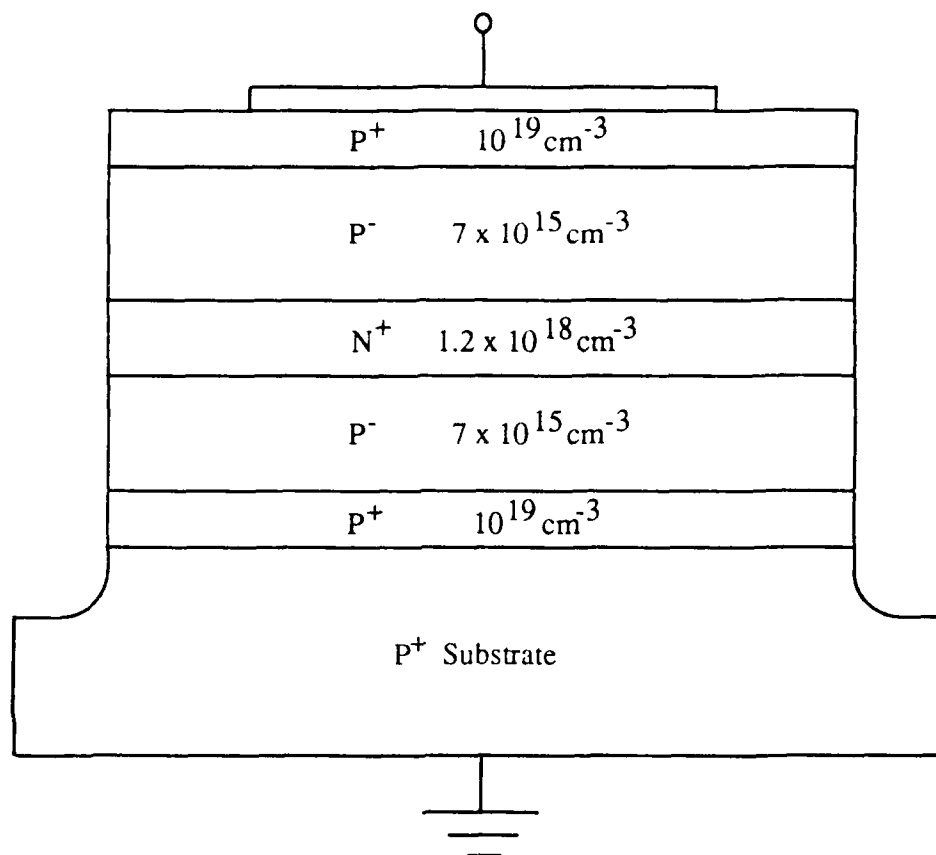


Figure 2.11 Experimental buried-well-capacitor test structure.

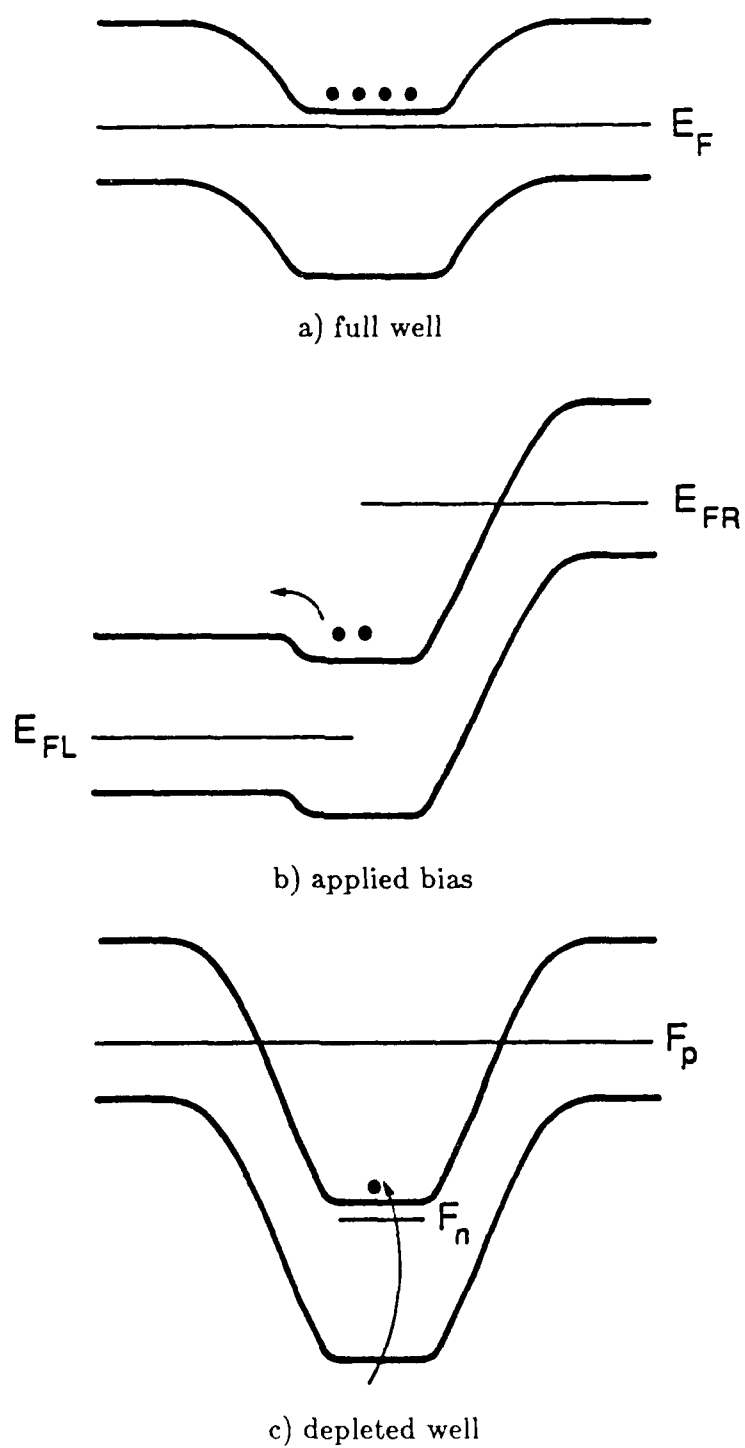


Figure 2.12 Energy-band diagram for the buried-well structure.

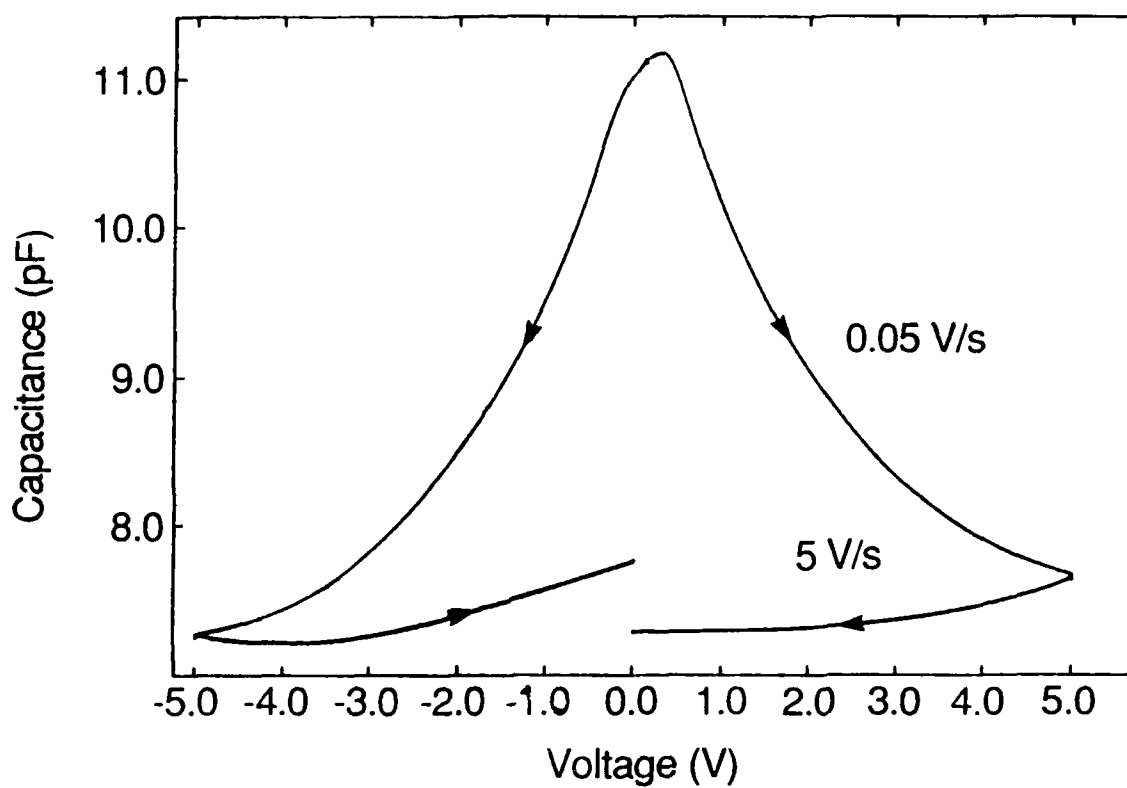


Figure 2.13 Capacitance-voltage curve for the buried-well structure at room temperature in the dark at 1MHz. Sweep rates from 0 to ± 5 V are 50 mV/s. Sweep rates on the return to zero bias are 5 V/s.

zero, one of the two PN junctions is slightly forward biased while the other is reverse biased, so the capacitance decreases. When the bias is returned rapidly to zero, the well remains depleted. Both junctions share the charge, and the capacitance remains unchanged. The capacitance will then slowly rise back toward its equilibrium value as generation replenishes the well and removes the reverse bias across the junctions. Observing the capacitance-recovery transient provides a method of measuring the generation rate in the device. A typical capacitance-recovery transient is pictured in Figure 2.14. The charge recovery is seen to be roughly exponential with time as predicted by equation 2.83. The capacitance-recovery time constant, τ_c , corresponds to the time required for the capacitance to recover $1 - 1/e$ (63.2%) of the difference between its depleted-well and full-well values. The measurement is made in the dark at 1 MHz, using a J-type thermocouple to monitor temperature.

The capacitance-recovery storage time is plotted versus inverse temperature in Figure 2.15. The initial bias used to dump the electrons from the well is 5V. At this bias neither the N^+ region or the P^- regions are completely depleted. The capacitance-recovery storage time of the cell is seen to decrease exponentially with temperature from approximately 180 seconds at room temperature down to 100 ms at 120 °C. The activation energy is defined by:

$$\tau_c = Ce^{E_A/k_B T} \quad [2.88]$$

where C is a temperature-independent proportionality constant and E_A is the activation energy, which is also assumed to be temperature independent. Then:

$$\log_{10} \tau_c = \frac{\ln \tau_c}{\ln 10} = m \frac{1000}{T} + b \quad [2.89]$$

where m and b are the slope and vertical intercept of the $\log_{10} \tau_c$ versus $1000/T$ plot. Using 2.88,

$$\ln \tau_c = \ln C e^{E_A/k_B T} = \frac{E_A}{k_B T} + \ln C = \frac{1000(\ln 10)m}{T} + (\ln 10)b \quad [2.90]$$

$$E_A = 1000k_B(\ln 10)m \quad [2.91]$$

From the slope of the line in Figure 2.15, the activation energy for the GaAs buried-well structure is 0.79 eV. Physically, the activation energy of a generation process is determined by the largest single energy transition an electron must make to move from valence band to conduction band. The activation energy of thermal generation through R-G centers located at mid-gap will go as $E_G/2$. To see this explicitly, note that equations 2.85, 2.86, and 2.87

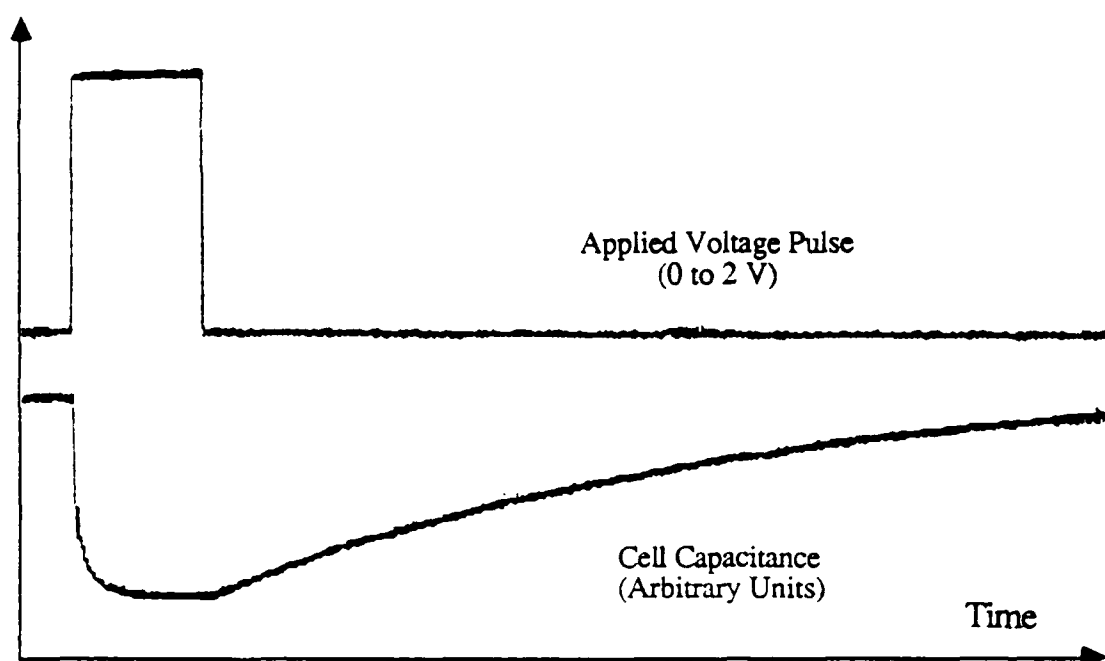


Figure 2.14 Capacitance-recovery transient for the buried-well test structure.

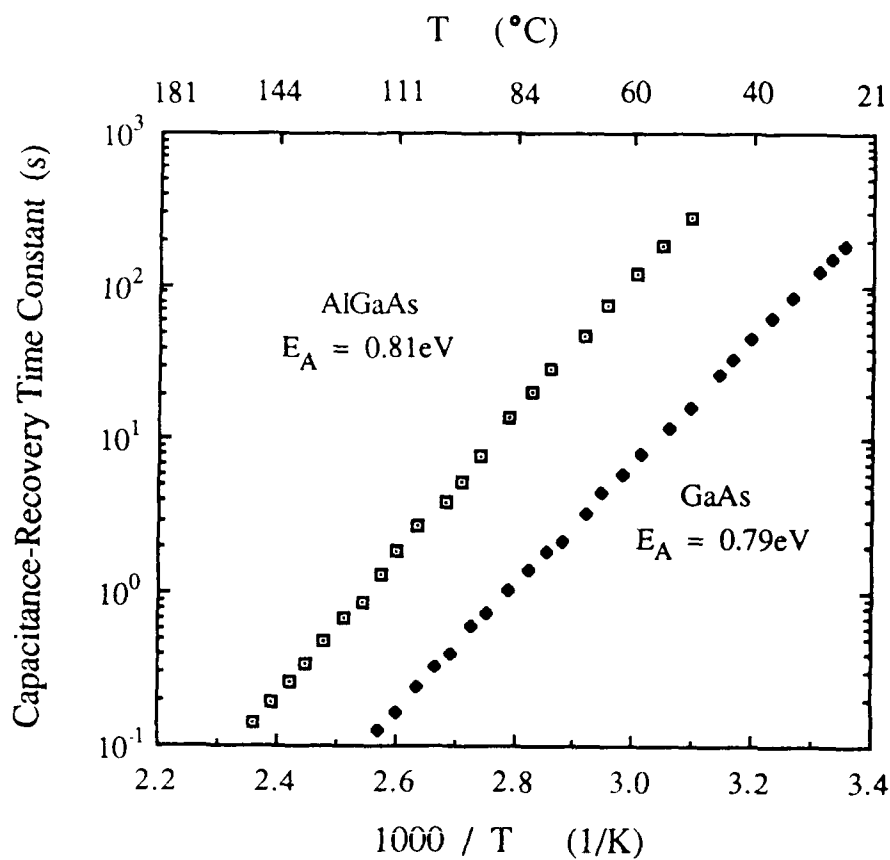


Figure 2.15 Temperature dependence of the 63.2%-capacitance-recovery time constant of buried-well structures in GaAs and AlGaAs.

can be used to express the storage time for bulk- or perimeter-limited diodes in the form of equation 2.88:

$$\tau_B = \frac{N_B \tau_G}{2\sqrt{N_C N_V}} e^{E_G/2k_B T} \quad [2.92]$$

$$\tau_P = \frac{N_B}{2s_0 \sqrt{N_C N_V}} \frac{A}{P} e^{E_G/2k_B T} \quad [2.93]$$

If the capture coefficients do not have strong temperature dependences, then equations 2.14 and 2.54 show that for generation due to R-G centers at the intrinsic energy level, τ_G and s_0 are approximately independent of temperature. The temperature dependence of the density of states terms is weak compared to that of the exponential. The bandgap varies approximately linearly with temperature over the range of temperatures in the measurement:

$$E_G = E_{G0} - \alpha T \quad (T > 300K) \quad [2.94]$$

where for GaAs, $E_{G0} = 1.58\text{eV}$ and $\alpha = 5.405 \times 10^{-4} \text{ eV/K}$. E_{G0} is not the zero-temperature bandgap of the material. It is instead the zero-temperature intercept of the linear approximation of $E_G(T)$. The storage time for a diode dominated by generation through bulk levels at mid-gap would then be:

$$\tau_B = \frac{N_B \tau_G}{2\sqrt{N_C N_V}} e^{\frac{E_{G0} - \alpha T}{2k_B T}} = \frac{N_B \tau_G}{2\sqrt{N_C N_V}} e^{-\alpha/2k_B} e^{E_{G0}/2k_B T} \quad [2.95]$$

$$= C_B e^{E_{G0}/2k_B T} \quad [2.96]$$

where:

$$C_B \equiv \frac{N_B \tau_G}{2\sqrt{N_C N_V}} e^{-\alpha/2k_B} \quad [2.97]$$

Similarly,

$$\tau_P = C_P e^{E_{G0}/2k_B T} \quad [2.98]$$

$$C_P \equiv \frac{N_B}{2\sqrt{N_C N_V}} \frac{A}{s_0 P} e^{-\alpha/2k_B} \quad [2.99]$$

Equations 2.96 and 2.98 show that the activation energy for generation dominated by mid-gap states would be $E_{G0}/2$, which matches the observed value precisely. Equations 2.90, 2.97, and 2.99 show that the vertical intercept of the $\log_{10} \tau_c$ -versus- $1000/T$ plot is related to the generation lifetime or the surface recombination velocity.

Because of the proportionality of the generation rates to the intrinsic carrier concentration, longer storage times should be achievable in wider-bandgap semiconductors, provided the surface generation velocity and generation lifetime do not change. To demonstrate this, the buried-well structure was also fabricated entirely in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ with an aluminum mole fraction of approximately 0.2. The dimensions and dopings were the same as those used in the GaAs buried-well structures. The AlGaAs devices exhibited over an order of magnitude increase in capacitance-recovery time constant, to 90 minutes at room temperature [35]. The time-constant-versus-temperature characteristic for an AlGaAs cell is plotted in Figure 2.15. The slope corresponds to an activation energy of 0.81 eV. The increased activation energy is due to the wider bandgap of the ternary compound.

The operating-temperature range of a dynamic memory cell with the storage-time performance observed in the buried-well test devices can be estimated by assuming that the maximum refresh cycle should be one hundredth of the storage time constant. Then, using a conservative high-speed dynamic memory refresh rate of 1 kHz, the cells could operate up to the temperature at which their storage time constants are reduced below 100 ms. From Figure 2.15 this temperature limit is seen to be near 120°C for the GaAs device, and near 150°C for the AlGaAs device.

2.2.4 Details of the Capacitance-Transient Measurement Technique

The simple derivation of the exponential charge recovery in a generation-limited PN-junction capacitor presented in Section 2.2.2.5 considered a single diode with symmetric doping (i.e., $N_A = N_D$). The experimental buried-well structures consist of back-to-back PN junctions in which the doping is highly asymmetric, with $N_D \gg N_A$. In the test structures described, the doping of the P regions above and below the N-type storage node is equal, but this is not mandatory, and in some cell designs it may be desirable to use different doping densities in the two P regions. The following subsections describe the differences in the capacitance transients of realistic buried-well structures compared to the simple exponential recovery predicted by equation 2.36 for the single symmetric diode.

2.2.4.1 Initial-Bias Dependence of the Storage Time

Expressions 2.85 and 2.86 predict that the charge-recovery time constant of a symmetrically-doped diode dominated by mid-gap traps will be independent of the magnitude of the initial stored charge. In the experimental structures, for which $N_D \gg N_A$, the capacitance-recovery time constant

exhibits variation with the magnitude of the initial bias used to remove electrons from the well. As illustrated in Figure 2.16, the capacitance-recovery time constant increases with increasing magnitude of bias pulses of either polarity. The capacitance-recovery time constant depends on the initial bias because the capacitance-recovery transient is not a true exponential. This occurs for two reasons. First, the capacitance is not directly proportional to the charge, so even if the charge recovery were truly exponential, the capacitance recovery would be slightly non-exponential. Second, in an asymmetrically-doped diode, the charge recovery itself is non-exponential. These two sources of time-constant dependence on initial bias are discussed in the following subsections.

2.2.4.1.1 Relation of Capacitance-Recovery Time Constant to Charge-Recovery Time Constant

For a buried-well structure with equal acceptor density in the P regions above and below the N region, the number of electrons per unit area missing from the well is given by:

$$Q = 2qN_D(x_n - x_{n0}) \quad [2.100]$$

$$Q = 2qN_D \left(\frac{N_A}{N_D + N_A} \right) (W - W_0) \quad [2.101]$$

where x_n , x_{n0} , W , and W_0 refer to depletion distances in either of the two identical junctions. The capacitance per unit area measured from the top contact to the substrate is the series combination of the two diode capacitances:

$$C = \frac{\epsilon}{2W} \quad [2.102]$$

In equation 2.102 it is assumed that the two junctions have equal areas and dielectric constants. To relate Q and C , solve 2.101 for W :

$$W = W_0 \left(\frac{Q}{2qN_D x_{n0}} + 1 \right) \quad [2.103]$$

Substituting 2.103 into 2.102 gives:

$$C = \frac{C_0}{1 + Q/Q_0} \quad [2.104]$$

where the constants C_0 and Q_0 are defined by:

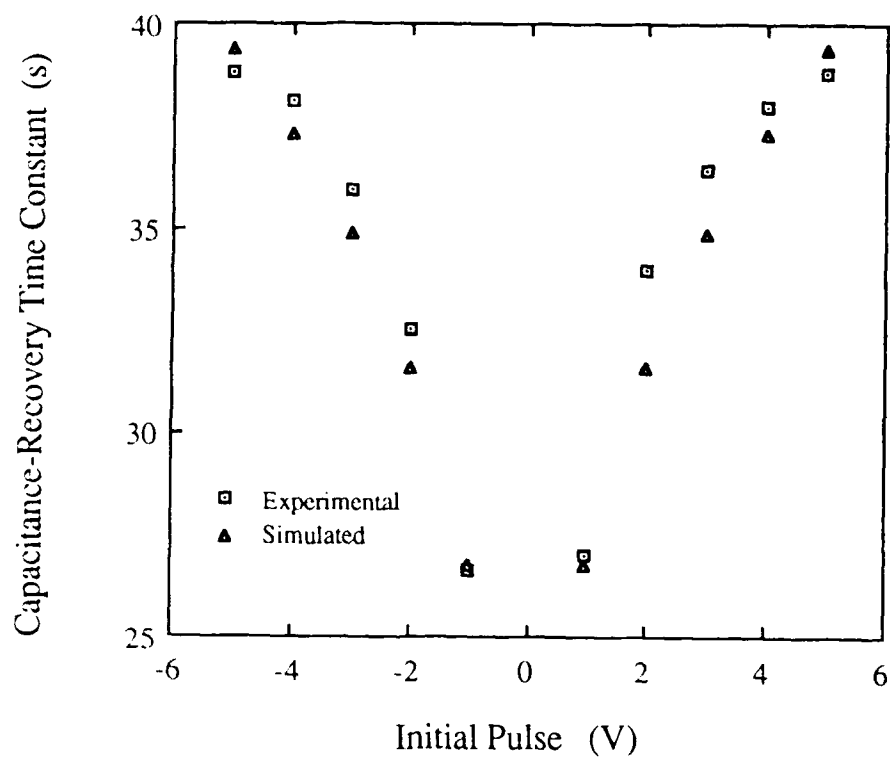


Figure 2.16 Initial-bias dependence of capacitance-recovery time constant. Squares indicate measured values; triangles indicate calculated values. Generation lifetime or surface recombination velocity was used as a fitting parameter. Data were taken at 41.5 °C.

$$C_0 = \frac{\epsilon}{2W_0} \quad [2.105]$$

$$Q_0 = 2qN_D x_{n0} \quad [2.106]$$

As described in Section 2.2.2.5, when $N_A = N_D$ and $E_T \cong E_i$, then the charge recovery will be an exponential, described by the charge-recovery time constant τ_S :

$$Q(t) = Q_I e^{-t/\tau_S} \quad [2.107]$$

Combining 2.107 and 2.104 gives:

$$C(t) = \frac{C_0}{1 + \frac{Q_I}{Q_0} e^{-t/\tau_S}} \quad [2.108]$$

The capacitance-recovery time constant, τ_c , is defined by:

$$C(\tau_c) = C_{\min} + (1 - 1/e)[C_{\max} - C_{\min}] \quad [2.109]$$

C_{\min} occurs at $t = 0$, and C_{\max} occurs at $t = \infty$, so:

$$C(\tau_c) = C_0 \left[\frac{1}{1 + Q_I/Q_0} + (1 - 1/e) \left(1 - \frac{1}{1 + Q_I/Q_0} \right) \right] \quad [2.110]$$

Finally, letting $t = \tau_c$ in equation 2.108 and solving simultaneously with 2.110 yields:

$$\frac{\tau_c}{\tau_S} = \ln \left[e \left(1 + \frac{Q_I}{Q_0} \right) - \frac{Q_I}{Q_0} \right] \quad [2.111]$$

The ratio of Q_I to Q_0 is determined by the magnitude of the initial bias pulse used to dump the well:

$$\frac{Q_I}{Q_0} = \frac{x_n(t = 0^+) - x_{n0}}{x_{n0}} \cong \left(1 + \frac{|V_P|}{4V_{bi}} \right)^{1/2} - 1 \quad [2.112]$$

$|V_P|$ is the magnitude of the initial voltage pulse. The factor of four arises because during the pulse, the initial voltage is essentially applied across the single reverse-bias junction, but during the transient, the two junctions are in parallel. The ratio of τ_c to τ_S is plotted versus initial bias magnitude in Figure 2.17. The figure is generated from equations 2.111 and 2.112 which assume that the charge-recovery transient is a true exponential.

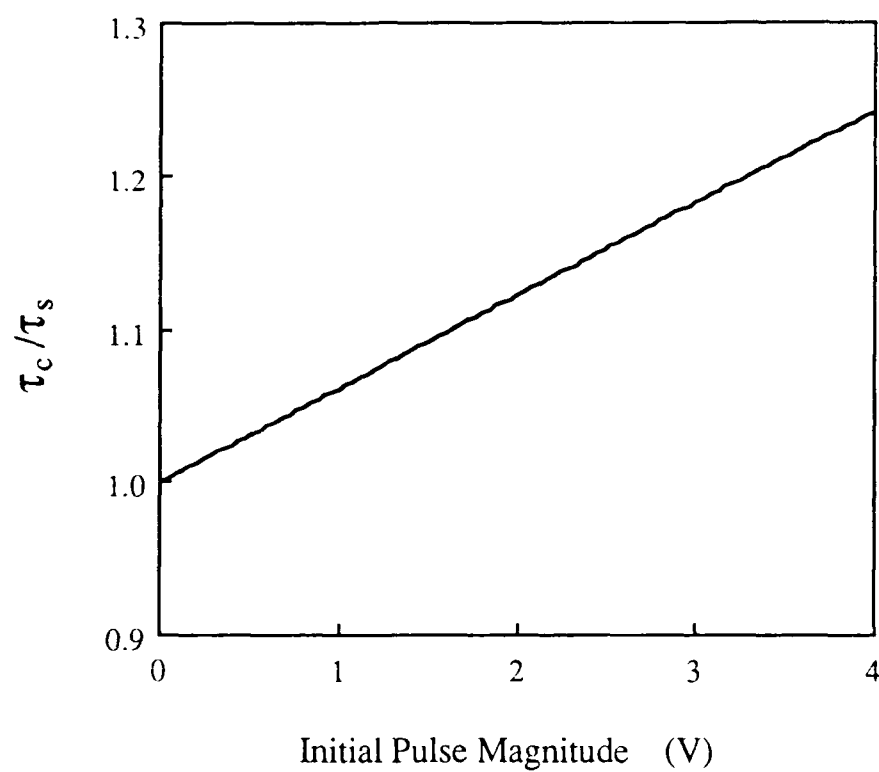


Figure 2.17 Ratio of capacitance-recovery time constant to charge-recovery time constant. In the calculation, the charge recovery is assumed to be exponential with time, and $N_A = N_D = 10^{17} \text{ cm}^{-3}$.

2.2.4.1.2 Non-Exponential Charge Recovery in Asymmetric Junctions

The charge recovery in asymmetrically-doped junctions ($N_A \neq N_D$) will be non-exponential because the generation width is not simply $W - W_0$ as assumed in equation 2.80 of Section 2.2.2.5. Instead, the generation width for the dominant trap levels is larger than $W - W_0$, and the ratio of W_G to $W - W_0$ decreases toward unity with increasing bias.

To derive an expression for the generation width as a function of doping and bias, consider the band diagram of an asymmetrically-doped diode in Figure 2.18. The generation width for a trap of energy E_T was defined in Section 2.2.2.1 as the region within which the trap energy level lies between the quasi-Fermi levels. The edges of the generation width are then defined by:

$$x'_p: E_T(x'_p) = F_P \quad [2.113]$$

$$x'_n: E_T(x'_n) = F_N \quad [2.114]$$

$$W_G(E_T) = |x'_p - x'_n| \quad [2.115]$$

The form of the expression for W_G is determined by the signs of x'_n and x'_p . There are three distinct cases: $x'_n > 0, x'_p > 0$ (W_G entirely on the P side); $x'_n < 0, x'_p < 0$ (W_G entirely on the N side); $x'_n < 0, x'_p > 0$ (W_G partially in each region). x'_n will be positive when the difference between the Fermi level and the trap level in the neutral N region exceeds the band bending on the N side of the junction. Define the voltage V_{TN} corresponding to the difference between the equilibrium Fermi energy on the N side and the trap energy level:

$$V_{TN} \equiv \frac{1}{q}(E_{FN} - E_T) = \frac{1}{q}[(E_{FN} - E_i) - (E_T - E_i)] \quad [2.116]$$

$$V_{TN} = \frac{1}{q} \left[kT \ln \left(\frac{N_D}{n_i} \right) - (E_T - E_i) \right] \quad [2.117]$$

For a specific trap level, V_{TN} is a constant as shown in Figure 2.18. For x'_n to be positive,

$$V_{TN} > \frac{N_A}{N_A + N_D}(V_{bi} + V_R) \quad [2.118]$$

where V_R is the reverse-bias voltage across the junction. So for $x'_n > 0$,

$$V_R < \frac{N_A + N_D}{N_A} V_{TN} - V_{bi} \quad [2.119]$$

Similarly, $x'_p < 0$ requires that:

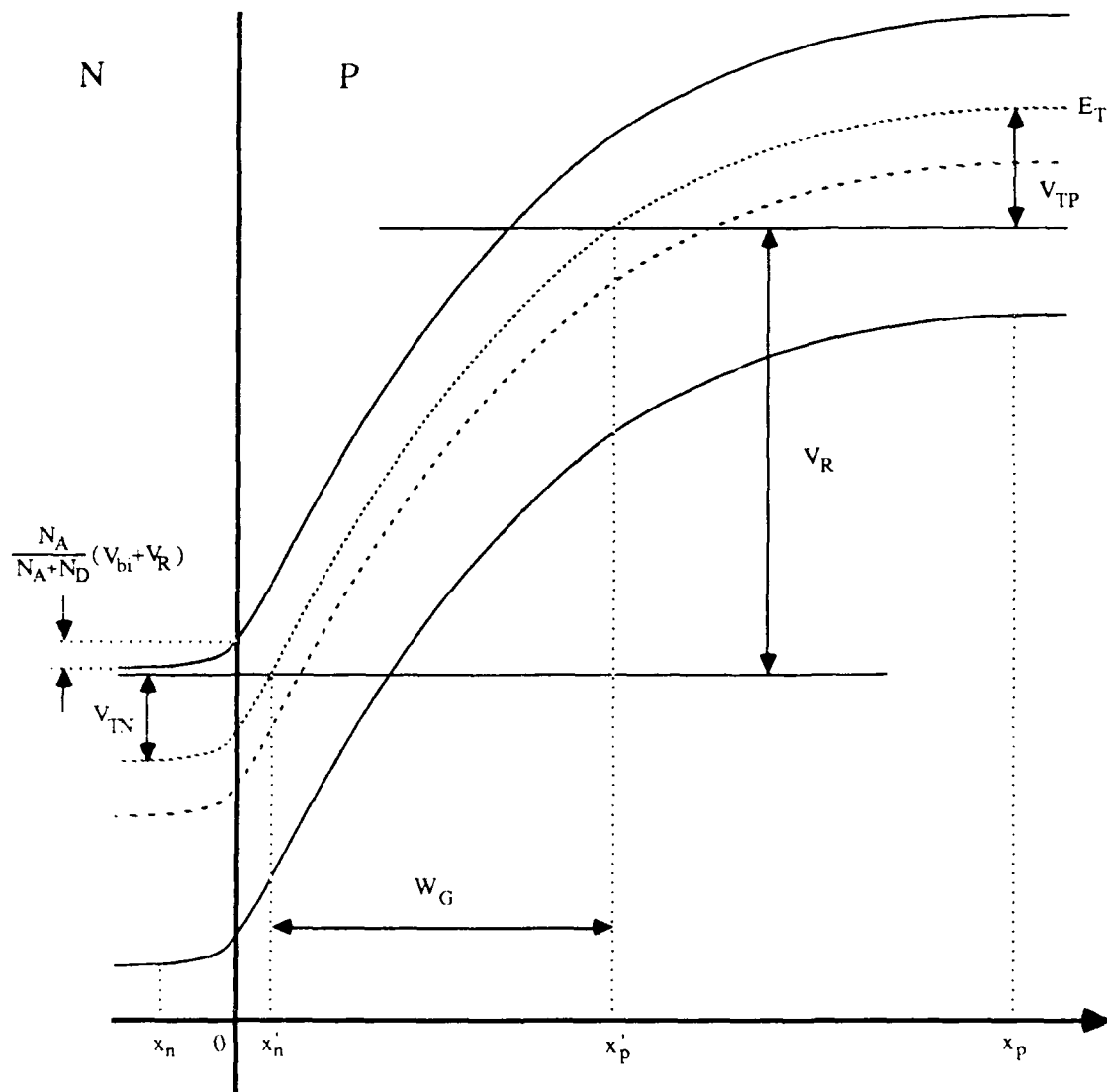


Figure 2.18 Generation width in an asymmetrically-doped junction.

$$V_R < \frac{N_A + N_D}{N_A} V_{TP} - V_{bi} \quad [2.120]$$

where V_{TP} is defined by:

$$V_{TP} = \frac{1}{q} \left[kT \ln \left(\frac{N_A}{n_i} \right) - (E_T - E_i) \right] \quad [2.121]$$

So for the generation width to lie entirely on the P side of the junction ($x'_n > 0$, $x'_p > 0$),

$$\frac{N_A + N_D}{N_D} V_{TP} - V_{bi} < V_R < \frac{N_A + N_D}{N_A} V_{TN} - V_{bi} \quad [2.122]$$

For the generation region to lie entirely on the N side of the junction ($x'_n < 0$, $x'_p < 0$),

$$\frac{N_A + N_D}{N_D} V_{TN} - V_{bi} < V_R < \frac{N_A + N_D}{N_A} V_{TP} - V_{bi} \quad [2.123]$$

When the generation width lies entirely on the P side, as is depicted in Figure 2.18 and described by equation 2.122, the generation width is given by:

$$W_G = \left[x_p^2 \left(1 + \frac{N_A}{N_D} \right) - \frac{2\epsilon}{qN_A} V_{TN} \right]^{1/2} - \left(\frac{2\epsilon}{qN_A} V_{TP} \right)^{1/2} \quad [2.124]$$

Similarly, when the generation width lies entirely on the N side under the condition of equation 2.123:

$$W_G = \left[x_n^2 \left(1 + \frac{N_D}{N_A} \right) - \frac{2\epsilon}{qN_D} V_{TP} \right]^{1/2} - \left(\frac{2\epsilon}{qN_D} V_{TN} \right)^{1/2} \quad [2.125]$$

Finally, when V_R exceeds both the limits defined by equations 2.119 and 2.120, the generation width extends across the metallurgical junction and,

$$W_G = W - \left(\frac{2\epsilon}{qN_A} V_{TP} \right)^{1/2} - \left(\frac{2\epsilon}{qN_D} V_{TN} \right)^{1/2} \quad [2.126]$$

Using $N_D = 10^{17} \text{ cm}^{-3}$ and $N_A = 10^{16} \text{ cm}^{-3}$, the ratio of the generation width calculated from equation 2.124 to $W - W_0$ is plotted in Figure 2.19 for three different trap levels. At low biases the ratio is larger. Equations 2.78 and 2.79 show that a larger ratio of W_G to $W - W_0$ corresponds to a larger ratio of leakage rate to stored charge. This is why smaller applied biases result in shorter storage times in asymmetrically-doped diodes, as shown in Figure 2.16.

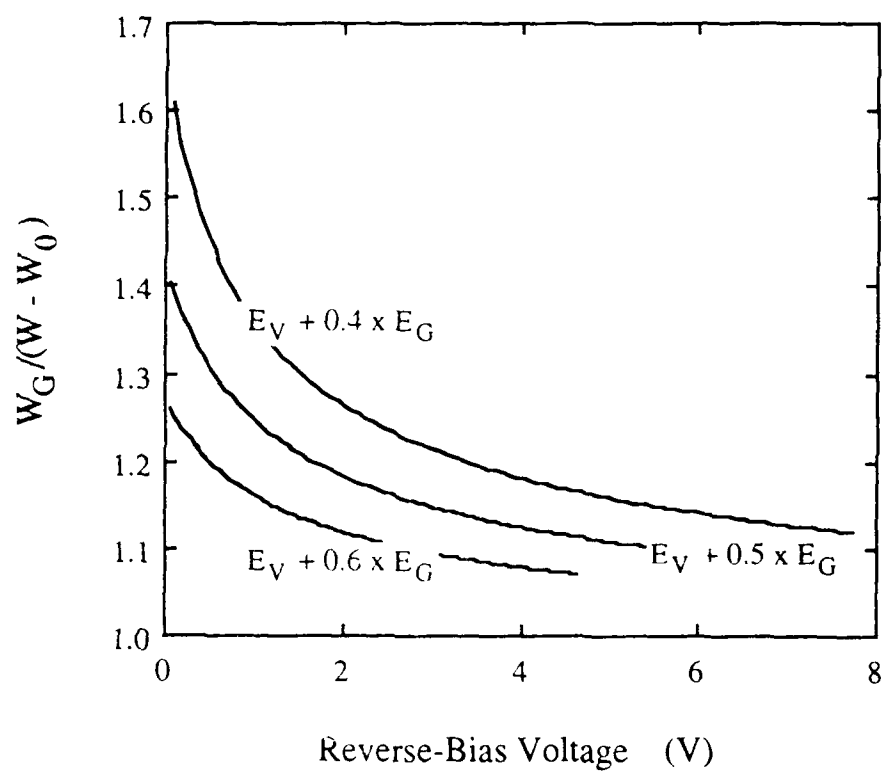


Figure 2.19 Ratio of generation width to $W - W_0$. The three curves represent trap levels at $E_V + 0.4 \times E_G$, $E_V + 0.5 \times E_G$, and $E_V + 0.6 \times E_G$. $N_D = 10^{17} \text{ cm}^{-3}$, $N_A = 10^{16} \text{ cm}^{-3}$.

When the generation width is not equal to $W - W_0$, the charge-recovery transient will not be a true exponential. The theoretical points in Figure 2.16 were calculated by starting at the initial bias, integrating the generation rate across the junction for a short time step, and then modifying the charge and bias appropriately. This procedure was repeated until the calculated capacitance of the structure recovered 63.2% of the difference between its initial and equilibrium values.

The equations in this subsection have been derived assuming the standard bulk diode band positions shown in Figure 2.18. At the perimeter of the device, the bands will be bent differently. However, as discussed in Subsection 2.2.2.3, the bands at the perimeter can be expected to have roughly the same form as the bands in the bulk within the perimeter generation width. In this case, equations 2.116 through 2.126 will describe devices dominated by generation in the bulk or at the perimeter.

2.2.4.2 Charge Redistribution During the Bias Pulse

In buried-well structures having different dopings in the upper and lower P regions, majority-carrier charge redistribution will result in instantaneous jumps in the capacitance when the initial bias pulse is applied and when it is removed. The total number of free electrons in the N-well varies only due to finite diode currents, so the total charge in the well does not change instantaneously. The applied biases across the two junctions must add to produce the bias applied across the whole device. Given a specific amount of charge in the well and an applied bias, the individual junction voltages can be uniquely determined. In a structure with different doping in the P regions, the back-to-back capacitors have different CV characteristics, and the total capacitance of the structure is determined by the voltage across the heavily-doped junction compared to the voltage across the lightly-doped junction. When the external applied-bias changes abruptly, the total charge initially stays the same, so the individual junction voltages must change as the electrons in the well shift from one junction toward the other. Such instantaneous changes in the junction voltages result in abrupt changes in the capacitance of structures with asymmetric doping in the P regions.

To relate the individual junction biases to the total applied bias and the net charge in the well, define N_E as the number of free electrons in the N region per cm^2 :

$$N_E = N_D(d - x_{n1} - x_{n2}) \quad [2.127]$$

where d is the depth of the N region and x_{n1} and x_{n2} are the depletion widths

into the N region from the upper and lower junctions, as shown in Figure 2.20. The figure is drawn with $N_{A1} > N_D > N_{A2}$. The depletion distances are related to the individual junction voltages by the usual expressions:

$$x_{n1} = \left(\frac{2\epsilon}{qN_D} \frac{N_{A1}}{N_D + N_{A1}} (V_{bi} - V_{A1}) \right)^{1/2} \quad [2.123]$$

$$x_{n2} = \left(\frac{2\epsilon}{qN_D} \frac{N_{A2}}{N_D + N_{A2}} (V_{bi} - V_{A2}) \right)^{1/2} \quad [2.129]$$

The individual junction applied voltages are related to the external bias by Kirchoff's rule:

$$V_A = V_{A1} - V_{A2} \quad [2.130]$$

where the polarity of V_{A1} and V_{A2} is from P to N, as usual. Together, equations 2.127 through 2.130 uniquely determine V_{A1} and V_{A2} for given N_E and V_A .

The instantaneous change in capacitance that occurs on sudden application of a bias is calculated as follows. Initially $V_A = V_{A1} = V_{A2} = 0$, and the equilibrium number of electrons in the well is calculated from 2.127. The initial capacitance is determined from the equilibrium depletion widths. Immediately after the bias is applied, N_E remains unchanged, and using the new V_A in equations 2.127 through 2.130 allows calculation of V_{A1} and V_{A2} . The new junction voltages determine the depletion widths, and thus the capacitance, immediately after application of the bias.

Using simple empirical expressions to model the diode current-voltage characteristics allows simulation of the charging of the device while the external bias is applied. The diode characteristics can be approximately modeled by:

$$J_D = J_{0f} e^{qV_D/n_f kT} \quad , \quad V_D > 0 \quad [2.131]$$

$$J_D = J_{0r} e^{-qV_D/n_r kT} \quad , \quad V_D < 0 \quad [2.132]$$

where J_{0f} , J_{0r} , n_f , and n_r are fitting parameters. The currents determine the change in the number of electrons in the well with respect to time:

$$J_{D1}(V_{A1}) + J_{D2}(V_{A2}) = -\frac{d}{dt}(qN_E) \quad [2.133]$$

Knowing the individual junction biases allows calculation of the currents, which modify N_E after a short time step. The new N_E allows recalculation of V_{A1} and V_{A2} , and so forth.

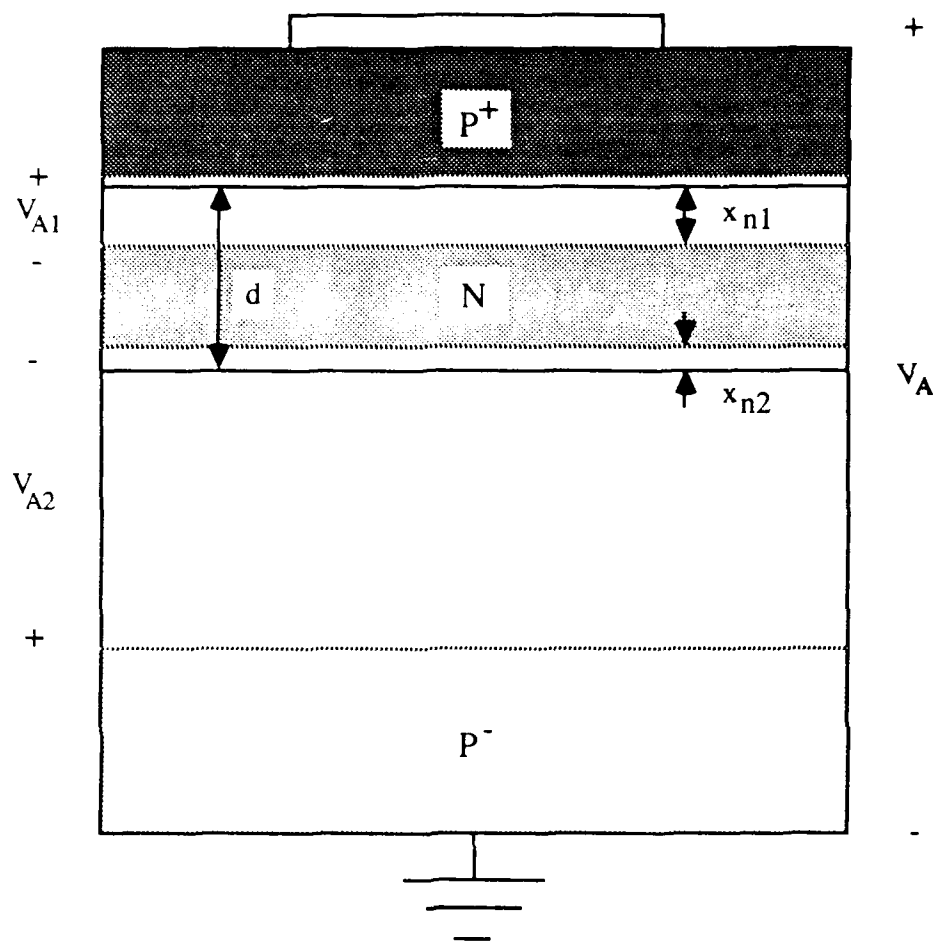


Figure 2.20 $P^+ - N - P^-$ buried-well structure.

Figures 2.21 and 2.22 show calculations of the capacitance, charge, and individual junction voltages during applications of positive and negative voltage pulses. The simulation of Figure 2.21 is for a +0.8V pulse of 1 second duration, while Figure 2.22 uses a pulse of -0.8V. The dopings used for the calculations are $N_{A1} = 10^{19} \text{ cm}^{-3}$, $N_D = 4 \times 10^{17} \text{ cm}^{-3}$, and $N_{A2} = 5 \times 10^{16} \text{ cm}^{-3}$. The fitting parameters are $J_{0f} = 5 \times 10^{11} \text{ A/cm}^2$, $J_{0r} = 8 \times 10^{-13} \text{ A/cm}^2$, $n_f = 2.0$, and $n_r = 2.6$. The area of the device is $1.33 \times 10^{-4} \text{ cm}^2$. Using these parameters, the simulated characteristics match those measured for real devices of the same area and doping very closely.

Application of a positive bias to the P^+-N-P^- structure causes the capacitance to jump to a lower value. This occurs because the lightly-doped junction's depletion region expands in reverse bias more than the heavily-doped junction's depletion contracts in forward bias. Application of a negative bias causes a positive jump in the capacitance because the heavier-doped region is expanding while the lighter-doped region contracts.

To use the capacitance-transient technique to measure time constants, it is essential that the initial capacitance is larger than the capacitance value immediately after the bias pulse ends. Under certain conditions with extremely asymmetric junctions like those simulated in Figures 2.21 and 2.22, it is even possible that application of pulses of one polarity will require additional electrons in the well. This effect is illustrated by the third plot of Figure 2.21, where the application of a positive voltage pulse produces a small negative charge in the well. After the positive pulse, the capacitance immediately returns to equilibrium, so no storage transient can be observed. Even for the negative pulse, a significant amount of time is required to charge the device.

2.2.5 Experimental Capacitor Performance Versus Doping Levels

Section 2.2.1 showed that the charge-storage density on a PN-junction capacitor at a fixed bias increases with increasing doping levels. Equations 2.85 and 2.86 of Section 2.2.2.5 show that if the generation lifetime is independent of doping, then the storage time will also increase with the doping levels. Both of these beneficial effects are simply due to decreasing depletion widths as doping increases. However, the built-in electric field also increases with doping level, resulting in field-enhanced generation [36,37], which decreases generation lifetimes in the bulk. Also, greater doping density may result in increased generation-site density which will also decrease generation lifetimes. Increasing doping levels will improve storage time performance as long as the beneficial effects of reduced generation volume outweigh the detrimental effects of shortened generation lifetimes.

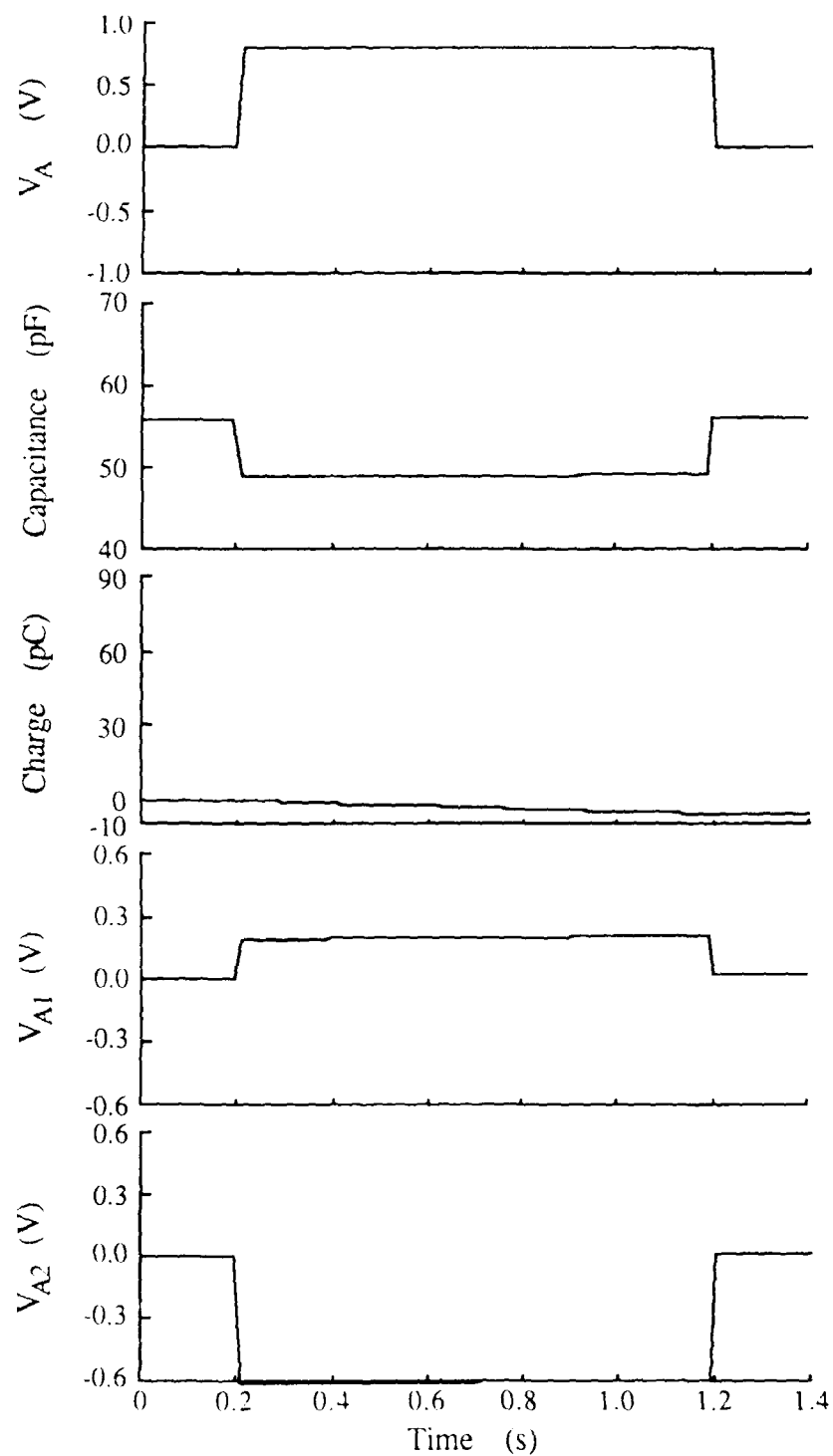


Figure 2.21 Simulation of asymmetric cell response to positive bias pulse.

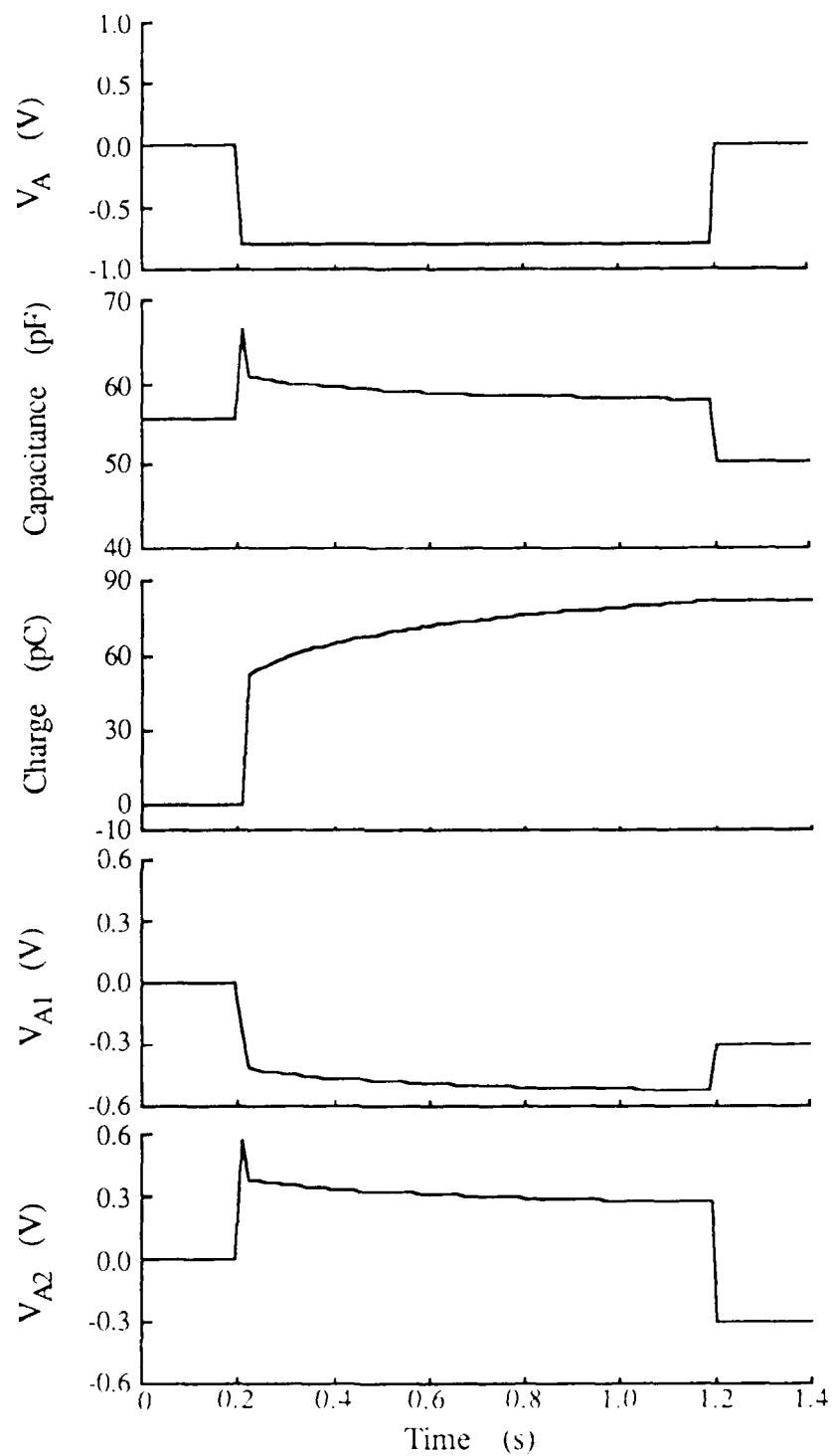


Figure 2.22 Simulation of asymmetric cell response to negative bias pulse.

Figure 2.23 shows storage-time-versus-temperature performance for three MBE-grown, etch-isolated PNP capacitor structures. The only significant differences in the structures are the doping in the P layers surrounding the floating N region. In all three structures, the doping in the N layer is 10^{18} cm^{-3} , while the P layers are doped at $7 \cdot 10^{15} \text{ cm}^{-3}$, 10^{17} cm^{-3} , and 10^{19} cm^{-3} in the three samples. All three devices thus have one-sided junctions, with the thermal generation occurring primarily in the lightly-doped material, as discussed in Section 2.2.4.1.2. In the two more-lightly-doped samples, the generation width lies in the P material, while in the sample with $N_A = 10^{19} \text{ cm}^{-3}$, the generation width is in the N type material.

It is apparent from Figure 2.23 that the competing beneficial and detrimental effects of increasing doping density result in a maximum in the storage time performance. The sample with the middle doping level produces the longest storage times. It is also apparent that increasing doping reduces the activation energies. Least-squares fits to the data give the activation energies as 0.79 eV, 0.62 eV, and 0.51 eV for the samples having $N_A = 7 \cdot 10^{15} \text{ cm}^{-3}$, 10^{17} cm^{-3} , and 10^{19} cm^{-3} , respectively.

A single activation energy is not a good description of the heavier-doped samples, because the $\log(\tau_t)$ versus $1/T$ characteristics are not strictly linear. The curvature in the characteristics suggests two different generation components with different temperature dependences. For example, the shape of the heavily-doped sample's curve can be explained by a bulk generation component with a low activation energy and a surface generation component with a higher activation energy. The bulk component tends to dominate at low temperatures, while the surface component dominates at high temperatures. Experimental support for this description will be presented in Section 2.2.6.

2.2.5.1 Field-Enhanced Generation

One of the reasons why the most-heavily-doped diodes do not produce the longest storage times despite having the smallest generation volume is that the higher fields associated with heavily-doped structures result in shortened generation lifetimes due to field-enhanced generation. In order for an electron to make its way from conduction band to valence band, it must first be excited into a generation center in the bandgap, and then escape from the trap into the conduction band. The electric field in the depletion region of the reverse-biased PN junction alters the shape of the potential barriers to electron and hole emission from generation centers. Figure 2.24 shows a Coulombic potential well with a barrier height E_{A0} at zero electric field, and with the barrier reduced by qE_A in the presence of an electric field.

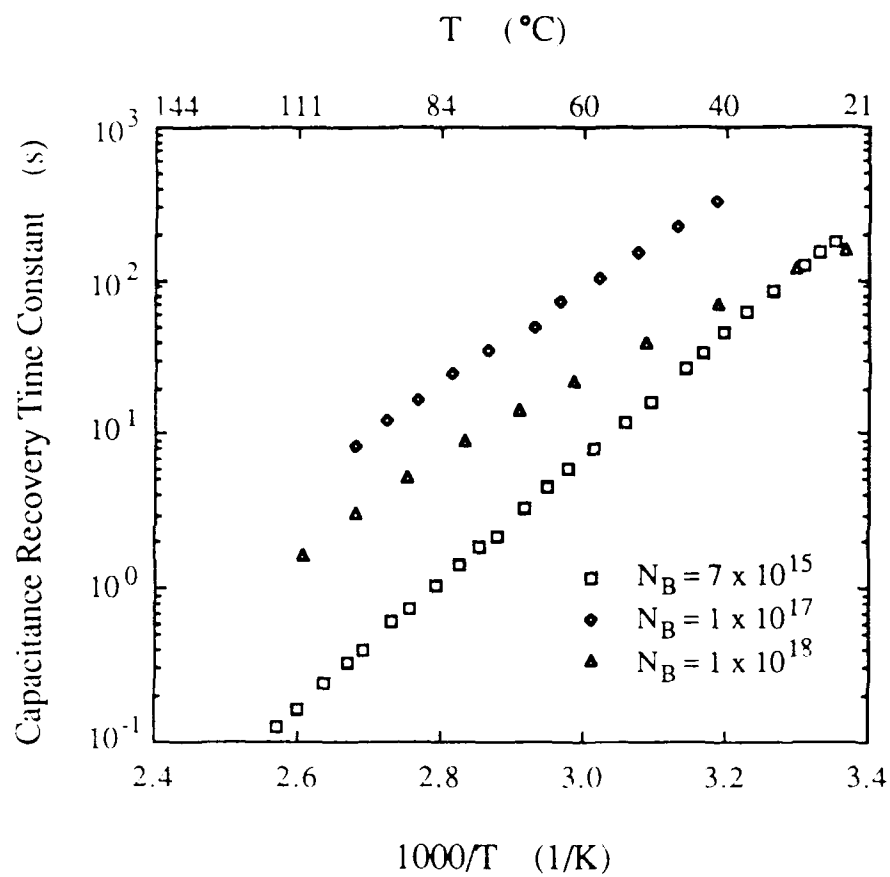


Figure 2.23 Capacitance-recovery time constant versus $1000/T$ for PN-junction capacitors with different doping levels.

The effects of an electric field on thermal generation were first described empirically by Poole [36], and theoretically by Frenkel [37]. Using the one-dimensional Coulombic potential of Figure 2.24, Frenkel derived a simple expression for the barrier reduction as a function of field strength:

$$\Delta E_A = 2q \left(\frac{q\mathcal{E}}{\pi\epsilon} \right)^{1/2} \quad [2.134]$$

where ΔE_A has units of eV, \mathcal{E} is the magnitude of the electric field in V/cm, and ϵ is the high-frequency dielectric constant of the material in F/cm. Figure 2.25 plots ΔE_A versus $\mathcal{E}^{1/2}$ from equation 2.134.

As described in Section 2.2.3, reducing the energy needed for thermal generation will reduce the activation energy derived from the storage-time-versus-temperature measurements. The peak electric fields at 0.25V reverse bias for the three samples from Figure 2.23 are 3.7×10^4 V/cm, 1.48×10^5 V/cm, and 5.2×10^5 V/cm for the diodes with $N_A = 7 \times 10^{15} \text{ cm}^{-3}$, 10^{17} cm^{-3} , and 10^{19} cm^{-3} , respectively. Figure 2.25 shows that the corresponding activation energy reductions expected from equation 2.134 for the three capacitors would be approximately 0.09 eV, 0.18 eV, and 0.33 eV. The experimental measurement does not provide ΔE_A directly, because the zero-field activation energy cannot be measured due to the built-in fields of the diodes. However, the differences between the observed activation energies for the different capacitors roughly correspond to the differences in the theoretical values of ΔE_A . This correspondence is somewhat unexpected, because previous attempts to match observed field-enhanced generation rates in silicon devices to the simple Frenkel theory have found the theory to greatly over-estimate the real effect [38-40].

It was pointed out by Hartke [41] that the potential barrier confining an electron at a generation site is three dimensional, and that equation 2.134 actually only describes the maximum directional barrier lowering, which occurs along the direction of the applied electric field. Evidence will be presented in Section 2.2.6 that perimeter generation is relatively unaffected by field enhancement. This could be explained by assuming that an electron trapped at a surface state must escape in a direction approximately normal to the surface. Section 2.2.2.3 concluded that there is very little electric field normal to the etched surfaces within the depletion region, so that field-enhanced barrier lowering for perimeter generation would be small and independent of doping.

For very large electric fields, it is also possible that the generation rate could be greatly increased by avalanche multiplication. In very-heavily-doped GaAs diodes, the built-in field approaches the critical field necessary for

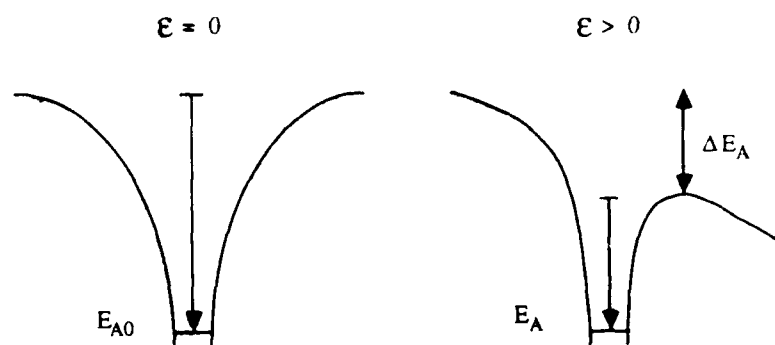


Figure 2.24 One-dimensional field-enhanced barrier reduction for a Coulombic potential.

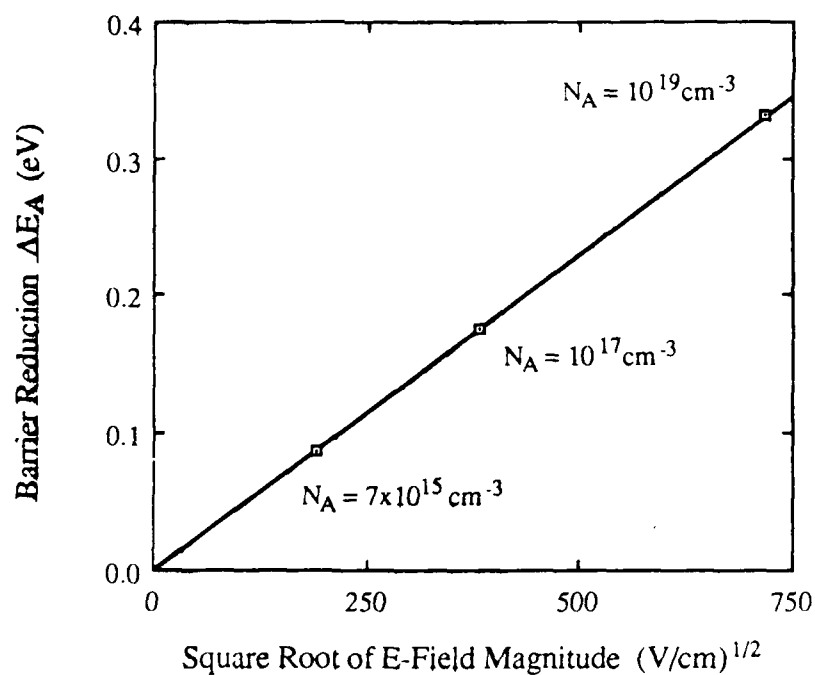


Figure 2.25 Calculation of field-enhanced barrier reduction versus square root of electric field magnitude.

avalanching to occur. However, for significant impact ionization, the critical field must exist over a distance sufficient to allow carriers to acquire the necessary kinetic energy. In other words, avalanching requires not only the critical field, but a minimum potential drop as well [42]. At the small biases which exist during the capacitance-recovery transients, and at the low GaAs logic levels used in a full dynamic memory, the required potential drop will not exist. Impact ionization should be negligible, unless it is somehow defect mediated so that the local energy band structure is altered.

2.2.5.2 The P^+iN^+ Capacitor

The calculated charge storage densities for the individual PN junctions shown in Figure 2.23 at one volt reverse bias are $0.19 \text{ fC}/\mu\text{m}^2$ for $N_A = 7 \times 10^{15} \text{ cm}^{-3}$, $0.68 \text{ fC}/\mu\text{m}^2$ for $N_A = 10^{17} \text{ cm}^{-3}$, and $2.11 \text{ fC}/\mu\text{m}^2$ for $N_A = 10^{19} \text{ cm}^{-3}$. High charge storage density is important for the design of dynamic memories with large single-chip capacity. The performance of the two more heavily-doped samples indicates that there is a design trade-off between storage time and charge storage capacity for the simple PN junction capacitor.

Long storage times and relatively high capacitance can be obtained simultaneously by inserting a thin undoped layer between the P and N regions of the diode. Figure 2.26 diagrams an experimental $P^+iN^+iP^+$ capacitor structure. The doping in the P^+ and N^+ layers is identical to the doping in the heavily-doped sample of Figure 2.23. The thickness of the undoped regions is 300\AA . The capacitance-recovery time constant measured for the $P^+iN^+iP^+$ structure is plotted versus temperature in Figure 2.27. For comparison, the performance characteristic of the $P^+N^+P^+$ structure is also included in the figure. The charge storage density for the individual P^+iN^+ junctions at one volt reverse bias is $1.85 \text{ fC}/\mu\text{m}^2$. The curvature of the characteristics will be discussed in Section 2.2.6.

Figure 2.27 shows that although the generation volume is smaller in the P^+N^+ devices, the storage time is longer in the junctions with the undoped layers. Because the doping in the P^+ and N^+ regions of the structures with and without the undoped layers is identical, the built-in potential of the two junctions must be the same. However, the maximum electric field is lower in the P^+iN^+ junctions, resulting in less field-enhanced generation. For example, the maximum field in the P^+N^+ junctions at 0.25V reverse bias is $5.2 \times 10^5 \text{ V/cm}$, while it is only $3.75 \times 10^5 \text{ V/cm}$ in the P^+iN^+ junctions. The one-dimensional theory of Figure 2.25 predicts that the trap-emission barriers will be reduced by approximately 50 mV in the P^+N^+ junctions compared to the P^+iN^+ junctions.

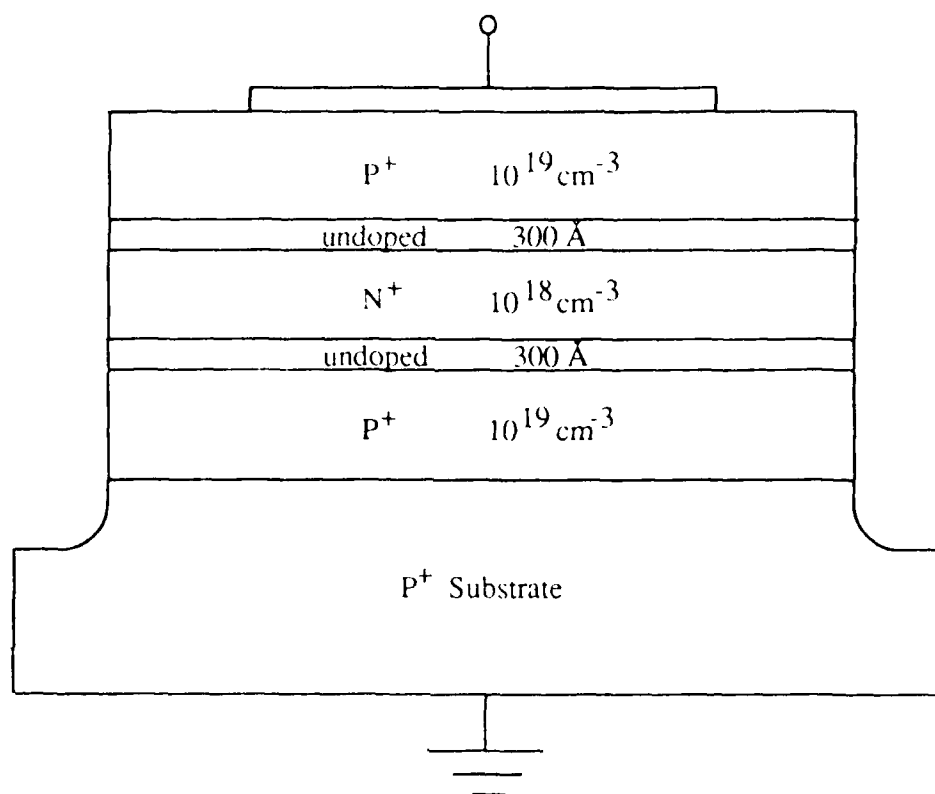


Figure 2.26 Experimental P⁺ iN⁺ iP⁺ capacitor.

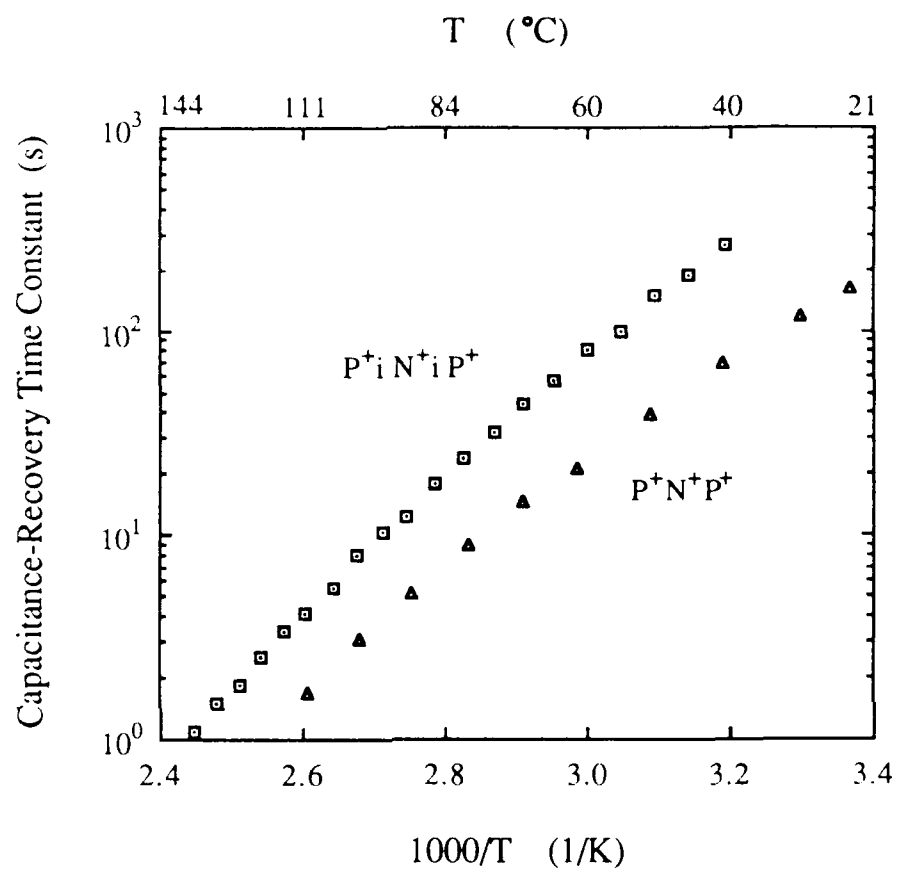


Figure 2.27 Comparison of storage-time performance of $P^+ i N^+ i P^+$ and $P^+ N^+ P^+$ devices.

Although the reduced electric field is probably the primary reason for the improved storage time of the P^+iN^+ device, there are two other factors which tend to improve its performance over the P^+N^+ device. The first is that the wider depletion width of the P^+iN^+ junction reduces the possibility of band-to-band tunneling. This is not a major effect, however, because tunneling-limited breakdown for ideal planar diodes with these doping levels is not expected to occur before the reverse bias reaches 5 volts [43]. Using a 1 volt dump pulse, the maximum bias across the junctions during the recovery transient is only 250 mV. The other feature of the P^+iN^+ junctions which may contribute to the improved storage time performance is that the generation width lies entirely in the undoped material. It is possible that the undoped material provides increased generation lifetimes compared to the degenerately-doped N^+ material within which generation occurs in the P^+N^+ structure.

With these doping levels, the 300Å undoped layers decreased the charge-storage density at 1 volt by a factor of about 1.14. The value of 300Å is somewhat arbitrary, and with empirical optimization the charge-storage reduction required to go from P^+N^+ to P^+iN^+ should be even less. Particularly if it is desirable to push the doping levels even higher than $N_A = 10^{19} \text{ cm}^{-3}$, $N_D = 10^{21} \text{ cm}^{-3}$, the P^+iN^+ structure should be beneficial.

2.2.6 Storage Time Dependence on Perimeter-to-Area Ratio

In a capacitor dominated by bulk generation, the storage time is independent of the area of the device, because the stored charge and the generation rate scale equally. In a device dominated by generation at the perimeter, the storage time will be inversely proportional to the perimeter-to-area ratio, because the charge scales with area while the generation rate scales with the perimeter. For a device with significant generation components due to the bulk and the perimeter, equation 2.84 from Section 2.2.2.5 shows that:

$$\frac{1}{\tau_s} = K \left[G_B + G_P \frac{P}{A} \right] \quad 2.135$$

where τ_s is the storage time, G_B is the bulk generation rate, G_P is the perimeter generation rate, P/A is the perimeter-to-area ratio, and K is a proportionality constant. The relationship expressed by equation 2.135 has been experimentally observed for the $P^+iN^+iP^+$ devices described in the previous section. Figure 2.28 plots inverse time constant versus perimeter-to-area ratio for several differently-sized $P^+iN^+iP^+$ capacitors at 130.5 °C.

The top axis in Figure 2.28 shows the edge length of a square capacitor which would correspond to the perimeter-to-area ratio on the bottom axis. The

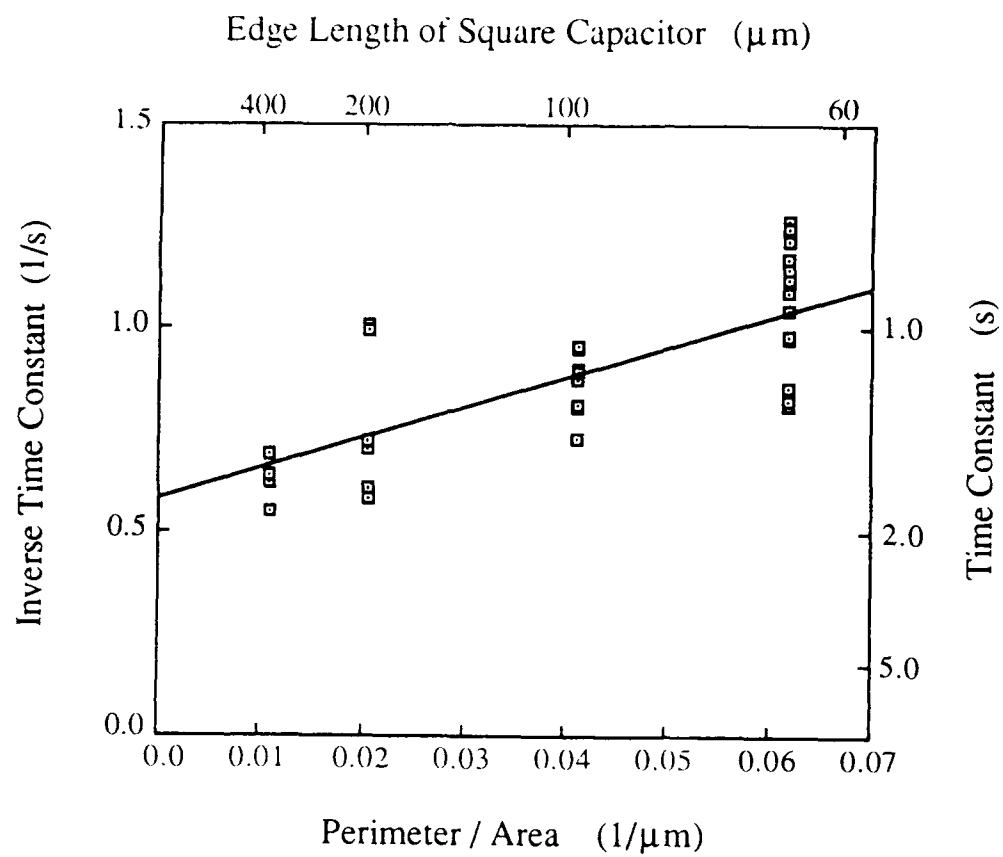


Figure 2.28 Inverse time constant versus perimeter-to-area ratio for $\text{P}^+ \text{iN}^+ \text{iP}^+$ capacitors at 130.5°C .

storage time is longest for the largest devices, and decreases as the devices become smaller and their perimeter-to-area ratio increases.

The dashed line represents a least-squares fit of the data. The non-zero vertical intercept of the line indicates the presence of a significant bulk generation component. The positive slope of the line indicates that there is also a significant perimeter generation component. Extrapolating the line off scale to the right of the figures allows prediction of the performance of very small area devices. For example, a $10 \times 10 \mu\text{m}$ device would have a perimeter-to-area ratio of $0.4 \mu\text{m}^{-1}$ and a storage time at 130.5°C of 730 ms. A $4 \times 4 \mu\text{m}$ device would have a perimeter-to-area ratio of $1.0 \mu\text{m}^{-1}$ and a 130.5°C storage time of 320 ms. These numbers indicate that if this performance can be achieved in a complete dynamic memory cell, then storage time reduction will not be a limit on minimum capacitor dimensions.

Storage-time-versus-temperature measurements for $\text{P}^+\text{iN}^+\text{iP}^+$ capacitors of different perimeter-to-area ratios suggest that the activation energies of the bulk and perimeter generation mechanisms are different. Figure 2.29 plots the capacitance recovery time constant for a large device and a small device. The perimeter-to-area ratios are $.011 \mu\text{m}^{-1}$ and $.062 \mu\text{m}^{-1}$ for the large and small capacitors, respectively. The large device has a longer storage time at all temperatures, and the difference clearly increases with temperature. Both characteristics curve slightly, but the activation energy of the smaller device is higher at each temperature. The smaller device with higher perimeter-to-area ratio is more perimeter dominated. The higher activation energy of the small device suggests that the activation energy of the perimeter generation component is higher than that of the bulk. The large electric field in these devices is responsible for the reduced activation energy of the bulk generation component. At sufficiently high temperatures, the slopes of both curves approach the 0.79 eV activation energy characteristic of generation through a continuous distribution of interface states, as described in Section 2.2.3. The activation energy of perimeter generation does not seem to be significantly reduced by field-enhanced generation, as discussed in Section 2.2.5.1.

2.2.7 Ion-Implanted Capacitors

All of the PN-junction capacitor structures described to this point have been fabricated by molecular beam epitaxy and isolated by wet etching. The capability to fabricate capacitors by ion implantation would be valuable to the design of complete memory cells, as will be described in the following chapters. All ion-implanted devices would be less costly than epitaxial designs, and planar processing is always desirable. The difficulty with ion implantation is that the

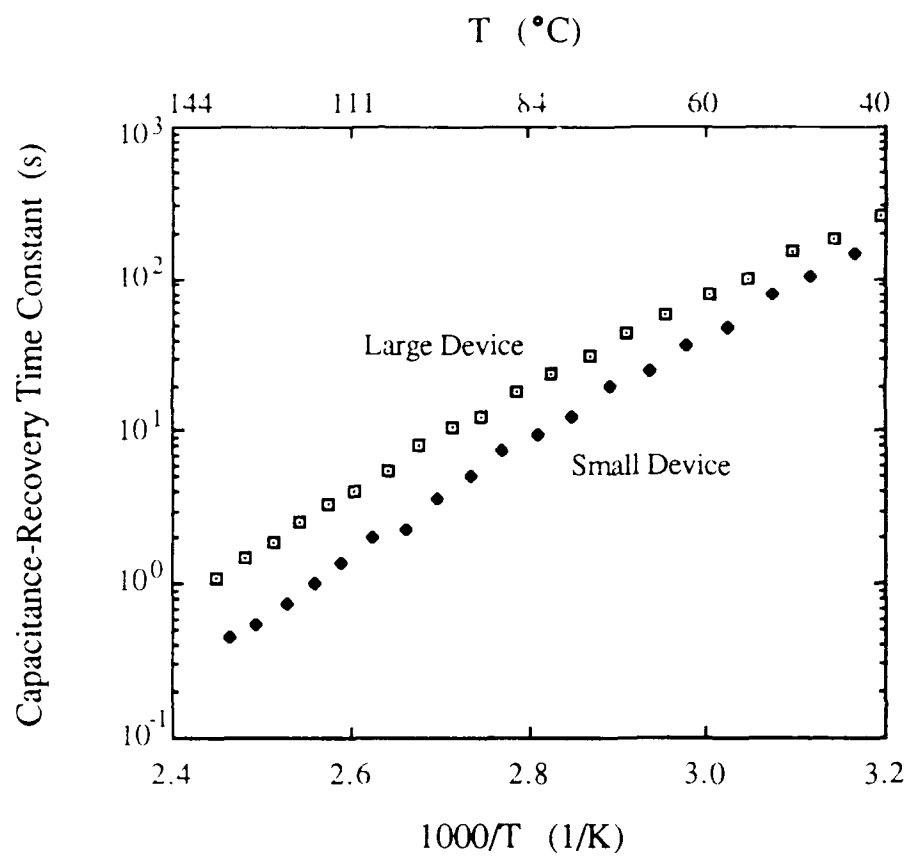


Figure 2.29 Comparison of storage time performance of large and small $P^+iN^+iP^+$ devices.

implantation is that the implantation produces lattice damage and interstitial atoms which will provide generation sites if not effectively annealed away.

2.2.7.1 Mg and Proton Implant-Isolated Devices

The first set of implantation experiments did not attempt to use implantation to form the PN junction of the storage capacitor. Instead, a shallow epitaxially-grown PNP structure was isolated by an unannealed proton bombardment and by an annealed magnesium implant, as indicated in Figure 2.30. Proton bombardment is a commonly used isolation technique in GaAs circuits which renders a region insulating by effectively destroying the crystal lattice [44]. Magnesium is a P-type dopant in GaAs. The proton bombardment used a dose of 10^{14} cm^{-2} at 45 keV, the magnesium implant used a dose of $2 \times 10^{14} \text{ cm}^{-2}$ at 200 keV (doubly ionized), and a rapid-thermal-anneal sequence of 10 seconds at 450°C and 12 seconds at 950°C in nitrogen gas. Both sets of capacitors were $300 \times 300 \mu\text{m}^2$.

The rapid-thermal anneal definitely activated the magnesium implant, because the upper and lower P^+ regions became electrically connected, but the activation percentage was not measured. The storage time in the Mg-isolated samples was measured by a current transient technique. A DC bias was applied from top to bottom contact causing a steady-state current. The bias was then pulsed momentarily to a higher value to extract electrons from the floating N region. The expanded depletion regions from the N layer caused the current to be reduced, and the recovery transient back to steady state revealed the storage time. The storage time of the proton-bombarded sample was measured by the standard capacitance transient technique. The time constants for the two samples are plotted versus temperature in Figure 2.31.

Both samples show greatly reduced storage times and activation energies compared to the epitaxial devices. The Mg-isolated devices show room temperature storage times of about 1 second, while the proton-bombarded devices show only about 100 ms of storage at room temperature. The activation energies are 0.13 eV for the Mg sample and 0.01 eV for the proton-bombarded sample. The low storage times and extremely low activation energies suggest a very high density of generation sites, and multiple trap interactions during the generation process. The results indicate that proton bombardment will probably not be a useful isolation technique for dynamic memory capacitors intended for use at room temperature and above. The magnesium-isolated device results suggest that capacitors formed or isolated by annealed implants are a realistic possibility, but that the implantation and annealing procedures would have to be improved.

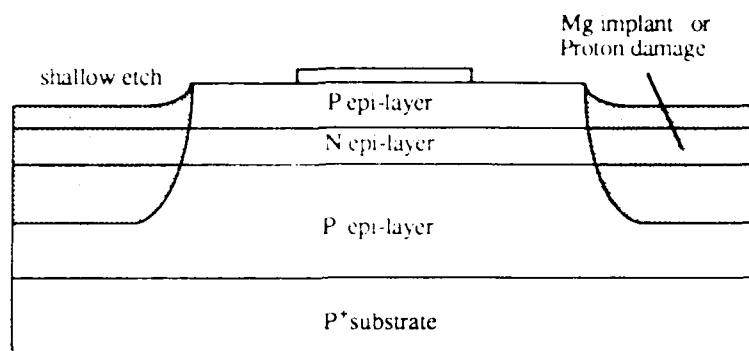


Figure 2.30 Implant-isolation test structure for initial Mg-implant and proton-bombardment experiments.

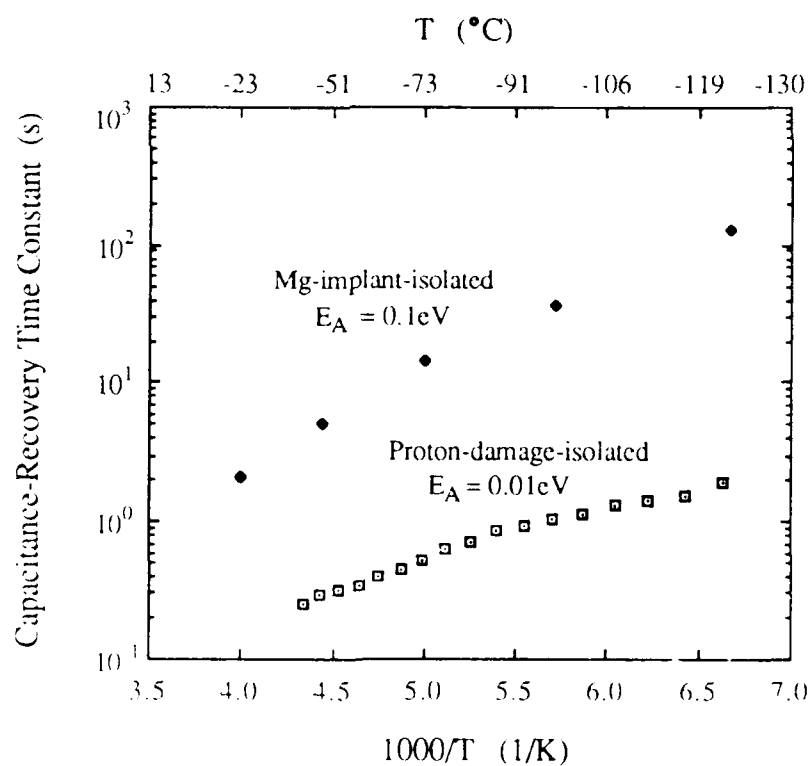


Figure 2.31 Recovery-time constant versus temperature for samples isolated by Mg implantation and by proton bombardment.

2.2.7.2 Capacitors Formed by Si and Mg Implantation

A second set of experiments has been performed in which an N-type storage region and an overlying P⁺ layer were implanted into a P-type epitaxial layer to form the planar PNP capacitor structure shown in Figure 2.32 [45,46]. The silicon implant was performed directly into the bare GaAs, while the Mg implant was performed through a thin CVD SiON layer. The SiON layer was used as an encapsulant during a single furnace anneal at 850 °C in N₂ with a GaAs proximity cap. The details of the fabrication and testing are reported in reference 46.

Combinations of various Si and Mg doses and energies have been examined. Figure 2.33 plots the capacitance-recovery time constant versus temperature for the best of the experimental devices. The capacitor measures 120×160 μm², and uses a dose of 5×10¹³ cm⁻² at 150 keV for the Si implant, and a dose of 5×10¹⁴ cm⁻² at 40 keV for the Mg implant. The background doping of the MBE-grown epilayer is 2×10¹⁵ cm⁻³. The cell's performance is greatly improved over the initial Mg-implant-isolation experiments. The storage time is 7 seconds at 300 K, and remains above 100 ms to over 100 °C. The activation energy is 0.52 eV, suggesting that the device is still bulk limited, so that it should not be severely affected by scaling.

The work of Loh et al. [47] indicates that further performance improvements may be possible through refinements in fabrication procedures. Their work reports diodes formed by Mg implantation into an N-type epilayer produced by VPE. The capacitance-recovery time is measured using a MIS/PN configuration, producing storage times of one minute at room temperature. Removal of the MIS structure indicates that even with these relatively long storage times, the structure was probably being limited by leakage due to the insulator rather than the implant. However, the generation width in their experimental structure lies primarily in the epitaxial layer and not in the implanted region. The result is encouraging for the fabrication of JFET access transistors by a combination of epitaxy and implantation as will be described in the next chapter. It is probable that continued experimentation will result in further improvements in the performance of ion-implanted PN-junction capacitors.

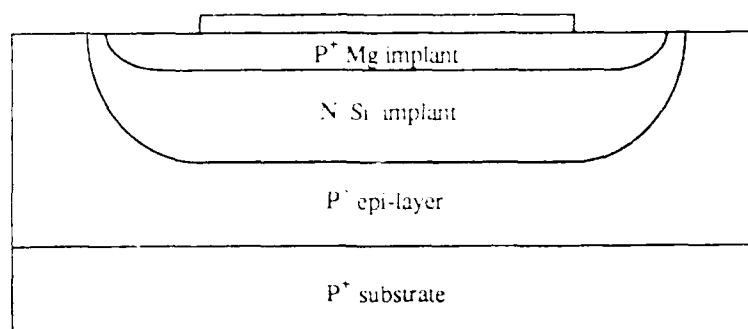


Figure 2.32 Planar PNP test structure formed by implantation of Si and Mg into a P-type epilayer.

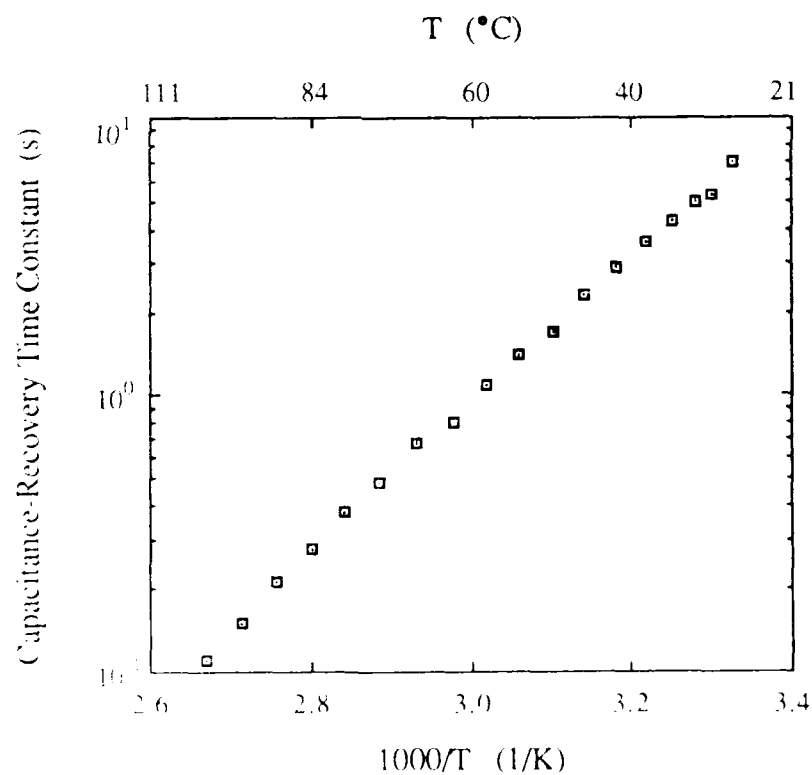


Figure 2.33 Capacitance-recovery time constant versus $1000/T$ for the planar Si and Mg implanted capacitor.

CHAPTER 3

ACCESS TRANSISTORS

3.1 Introduction

To make a dynamic memory cell, it is necessary to have a way of moving charge on and off of the storage capacitor. There also must be a way of selecting a single bit out of a two-dimensional array of cells. In silicon designs, an insulated-gate field-effect transistor provides both functions. As diagramed in Figure 1.2, the gate of the transistor is connected to the wordline, which selects the cell, while the source is connected to the bitline, which moves charge on and off the capacitor. The drain of the transistor is directly connected to the storage node, so that the leakage current from the transistor gate and source will be collected by the capacitor in the storage state. As discussed in Section 2.1, MOS transistors are not available in GaAs. There are a variety of other types of GaAs field-effect transistors to be considered as candidates for the access elements in the dynamic memory cell.

It is conceivable that a buried storage node might have no direct electrical connection to the access transistor. It might instead be punch-through isolated, having an undepleted layer of P-type material separating the N-well from the N-type drain in the storage state. To access a punch-through isolated cell, the drain of the access transistor is raised to some positive potential to cause the drain depletion region to touch that of the buried well so that charge sharing occurs. The advantage of punch-through isolation is that the access transistor adds no additional storage-time-limiting leakage currents to the cell. The disadvantages include complicated cell design, fabrication, and operation, and clumsy selection of a single cell. Because of these disadvantages, punch-through isolation will not be considered further.

Bipolar transistor access to PN junction storage capacitors is similar to punch-through access, except that contact is made to the P region that separates the storage node from the N region of the transistor. The P region becomes the base of the bipolar access transistor, and the storage node acts like a floating collector. This is an attractive possibility because it may provide the low

leakage of punch-through isolation with the ease of operation of a directly-connected design. However, selectively reading the cell is dependent upon satisfactory gain in the transistor, which is difficult to achieve at low current levels in GaAs bipolar transistors. Bipolar access is a promising research topic, but it will not be discussed further in this work.

Modulation-doped field-effect transistors (MODFETs) also offer some attractive features as access transistors. The speed performance of MODFETs increases drastically with decreasing temperature, making a MODFET the obvious access-transistor choice for a memory designed for operation at cryogenic temperatures, where PN-junction capacitors are essentially static. Even at room temperature, a modulation-doped storage capacitor might offer improved storage-time performance over the PN-homojunction cells. This is because a modulation-doped storage region might be isolated by an enhancement-type ring, requiring no etched perimeter through the junction. MODFETs may also provide lower drain-to-source leakage in the "off" state than is achieved by traditional doped-channel FETs. In the "off" state the MODFET's two-dimensional electron gas is eliminated, and a lightly-doped P region separates the N-type source and drain. The P region provides a larger electrostatic barrier between source and drain than would be present at the same gate bias in a transistor with an N-doped channel. MODFET gate leakages depend strongly on the design of the gate, and are presently being evaluated. MODFET-accessed dynamic memory cells are currently under research at Purdue University, but they will not be discussed further in this work.

In the absence of a gate insulator, field-effect transistors in GaAs are traditionally formed by using a PN junction (JFET) or a metal-semiconductor junction (MESFET) to gate an N-doped channel between N^+ source and drain regions. When such a transistor is directly connected to a PN-junction capacitor to form a dynamic memory cell, there are two critical requirements for the transistor's performance. The first requirement is that the drain leakage current in the "off" state must be small enough to allow reasonable storage times over the desired range of operating temperatures. The second is that the gate current in the "on" state must be small enough so that it does not mask the signal charge coming from the capacitor during a read operation. Secondary performance criteria require the transistor to move the charge on and off the capacitor at high speed, and to dissipate very little power in the storage state. The remainder of this chapter will describe how MESFET and JFET design parameters such as dimensions and doping are related to these access-transistor performance requirements.

3.2 Subthreshold Leakage

The storage time in a PN-junction-capacitor-based dynamic memory cell is determined by the length of time required for leakage currents to transform a non-equilibrium logic "1" into a logic "0". Leakage currents due to thermal generation in the capacitor itself were described in the previous chapter. This section describes the additional leakage currents introduced by the addition of a directly-connected access transistor. A logic "1" is produced by removing electrons from the capacitor so that the storage node, connected to the transistor's drain, is positively charged with respect to the grounded substrate. In the storage state the transistor's drain leakage currents will discharge the capacitor. The objective in designing a direct-accessed cell is to select the transistor parameters and the voltages in the storage state to minimize the drain leakage current.

The typical dependence of drain leakage current on gate-to-source bias for several types of doped-channel FETs is shown in Figure 3.1. This figure suggests two primary regions of operation below threshold. In the first region the drain current consists of leakage to the source, which decreases exponentially with increasing reverse bias on the gate. At sufficiently negative gate biases, the reverse-bias leakage of the gate becomes dominant, and the drain leakage slowly increases with further reverse bias applied to the gate. The transistor should be designed so that the minimum drain leakage current is as low as possible, and the circuit must be designed so that the transistor is at its minimum leakage point during the storage state.

The gate leakage component should decrease with increasing gate-to-channel barrier height. This is suggested by the four hypothetical transistor leakage characteristics shown in the figure. The four gate leakage curves are drawn for different types of doped-channel FETs with increasing barrier heights. A standard GaAs MESFET provides a barrier of about 0.7 eV, which is independent of the workfunction of the gate material due to surface Fermi-level pinning. The second characteristic is for a MESFET with the GaAs surface treated with an ammonium-sulfide solution just prior to gate-metal deposition. This technique modifies the surface state density and makes the metal-semiconductor barrier height sensitive to the metal workfunction. Schottky barriers of greater than 0.9 eV have been demonstrated using gold on N-type GaAs [19]. As suggested in the figure, MESFETs with treated gates show reduced gate leakage compared to standard MESFETs.

The third characteristic shown is for a conventional GaAs JFET. The PN-junction gate of the JFET eliminates the thermionic emission which dominates

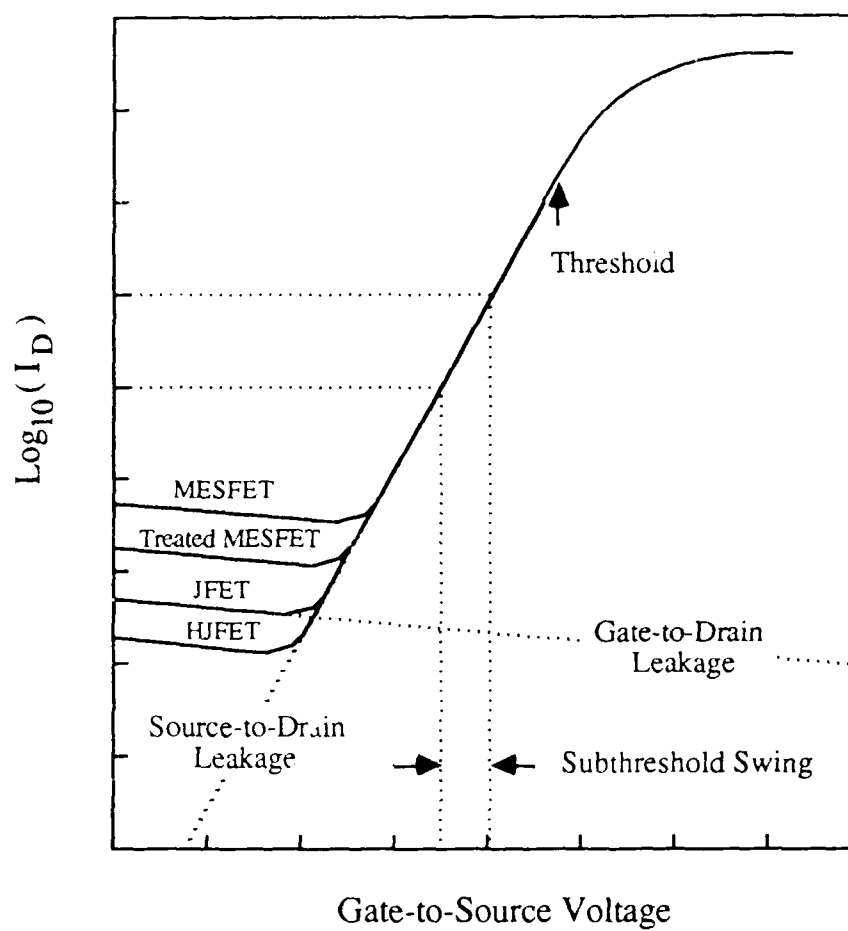


Figure 3.1 Conceptual plot of the logarithm of drain current versus gate-to-source voltage for four types of field-effect transistors.

the MESFET. The JFET leakage current is due to the bulk and perimeter generation currents of the reverse-biased diode, as described in the previous chapter. The final characteristic suggests that even lower gate leakage might be obtained by using a heterojunction JFET as the access transistor. Using a wider bandgap material for the P region of the gate would help to suppress gate leakage, just as it helps to increase the storage time in PN-junction capacitors.

The gate leakage component is critical to the determination of the minimum drain current. However, for simple MESFETs and JFETs, the dependence of the leakage current on physically controllable parameters is either obvious (e.g., thermionic emission dependence on barrier height) or very difficult to model (e.g., PN-junction perimeter generation rate dependence on doping concentration). The gate leakage current density can be treated as an empirical parameter characteristic of the type of access transistor selected. As illustrated in Figure 3.1, the gate leakage current is only weakly dependent on gate voltage. Doped-channel FET gate leakages are commonly modeled by [48]:

$$I_{DG} = G_{DG0} V_{DG} e^{-\frac{qV_{DG}}{kT}} \quad [3.1]$$

where G_{DG0} and δ are empirical fitting parameters.

The drain-to-source leakage current decreases exponentially with increasing reverse bias on the gate, producing a straight line on the logarithmic plot. The most important parameter describing the source leakage current is the subthreshold slope, the slope of the line on the $\log(I_D)$ versus V_{GS} plot. To reduce the reverse-bias voltage which must be applied to the gate to reach the minimum drain current, the subthreshold slope should be as large as possible. For drain-to-source biases larger than a few kT/q , the drain-to-source leakage current is only dependent on drain-to-source voltage through the dependence of the threshold voltage on drain-to-source voltage. Drain-to-source leakage current can be modeled by [49]:

$$I_{DS} = I_{DS0} \exp\left(\frac{q}{n_s kT}(V_{GS} - V_T)\right) \left(1 - \exp\left(-\frac{qV_{DS}}{kT}\right)\right) \quad [3.2]$$

$$V_T = V_{T0} + \kappa V_{DS} \quad [3.3]$$

In equations 3.2 and 3.3, I_{DS0} is an empirical fitting parameter, n_s is the subthreshold ideality factor, V_{T0} is the threshold voltage extrapolated to zero drain-to-source voltage, and κ is a negative constant describing the dependence of threshold voltage on V_{DS} .

By assuming that the drain-to-source leakage current varies exponentially with the minimum potential barrier height between source and drain, it is

possible to derive n_s , V_{T0} , and κ from two-dimensional calculations of the potential in the subthreshold transistor. The next subsection describes a modeling technique for performing a two-dimensional calculation of the potential in subthreshold MESFETs and JFETs on lightly-doped substrates. The potential solutions allow calculation of the drain-to-source leakage current parameters as a function of the physical design parameters, such as transistor dimensions, dopings, and voltages.

3.2.1 Harmonic Solution for Potential in Subthreshold FETs

Calculating the potential in a subthreshold FET requires solution of the two-dimensional Poisson equation in the plane defined by the normal to the semiconductor surface and a line along the channel between source and drain. The most direct approach to solving the two-dimensional Poisson equation is an iterative solution to a finite-difference formulation using successive-over-relaxation. In a region in which the charge density is a function of only one variable, such as in the depletion region under the gate of a FET, the two-dimensional Poisson equation can instead be treated as the sum of a one-dimensional Poisson equation and a two-dimensional Laplace equation:

$$\text{2D Poisson: } \nabla^2 \phi(x,y) = \frac{-\rho(y)}{\epsilon}, \quad \text{becomes}$$

$$\text{2D Laplace: } \nabla^2 \psi(x,y) = 0, \quad \text{and}$$

$$\text{1D Poisson: } \frac{d^2}{dy^2} U(y) = \frac{-\rho(y)}{\epsilon}, \quad \text{with}$$

$$\phi = U + \psi$$

ϕ is the two-dimensional potential, ρ is the charge density, and ϵ is the dielectric constant. ψ is the solution of the two-dimensional Laplace equation, and U is the solution of the one-dimensional Poisson equation in the direction in which the charge density is varying (i.e., normal to the semiconductor surface in a FET). The heart of the solution is choosing appropriate boundaries and boundary conditions for the 2D Laplace equation.

Marshall and Meindl [50] applied this approach to Si MESFETs on doped substrates. On a doped substrate, the current is sufficiently confined to the channel region so that some error in the boundary conditions below the channel does not significantly affect current predictions. Accordingly, Marshall and Meindl produced accurate agreement with a full numerical solution by using 1D

Poisson equations through the source and drain as the boundary conditions perpendicular to the channel at the source and drain ends of the gate. Their work specifically excluded GaAs MESFETs, because on an undoped substrate the fields far below the channel can become significant.

Using the one-dimensional source and drain solutions as boundary conditions for the solution of the Laplace equation in the region below the gate is equivalent to assuming that the gate has no effect on the potentials below the drain and source. This results in a discontinuous derivative of the potential along the gate-to-source and gate-to-drain boundaries, as can be seen in Figure 3.3a. In this work, the use of the 1D Poisson solutions as boundary conditions for the 2D Laplace equation is avoided by solving the 2D Laplace equation separately under the source, gate, and drain. The three regions of the solution are indicated in Figure 3.2. The source and drain regions are considered to extend to infinity in the x direction of the figure. The Fourier coefficients for the solution to the 2D Laplace equation in the three regions can be uniquely determined by requiring continuity of the potential and its normal derivative along the common boundaries.

The charge density in the source and drain regions indicated in the figure is not strictly a function of only one variable as required for the separation of the 2D Poisson equation into a 1D Poisson equation and a 2D Laplace equation. **There will be some depletion of the N-doped layer of the source and drain adjacent to the gate.** In this work, this depletion is ignored: the charge density in the N-layer is treated as changing abruptly from the channel doping to zero at the gate-to-source and gate-to-drain boundaries.

For a harmonic solution by separation of coordinates in a Cartesian system, the substrate boundary condition must be imposed along a line parallel to the gate. Consider, for example, the back boundary to be selected as $y = y_B$. The depletion edge of the real solution must be a function of x , so it is impossible to select y_B such that the area of the solution remains entirely within the depletion region and $\phi(x, y_B) \neq f(x)$. If the potential were known along some line $y = y_B$ within the depletion region, then the back boundary condition could be included in the solution through a Fourier series of terms which are harmonic in the x direction of Figure 3.2. However, because the potential $\phi(x, y_B)$ is not known, the most convenient approximation is to require $\phi(x, y_B)$ to be a constant, ϕ_{BB} . Then the back boundary condition for the Laplace solution becomes $\psi(x, y_B) = 0$, and the one-dimensional Poisson equation must satisfy $U(y_B) = \phi_{BB}$.

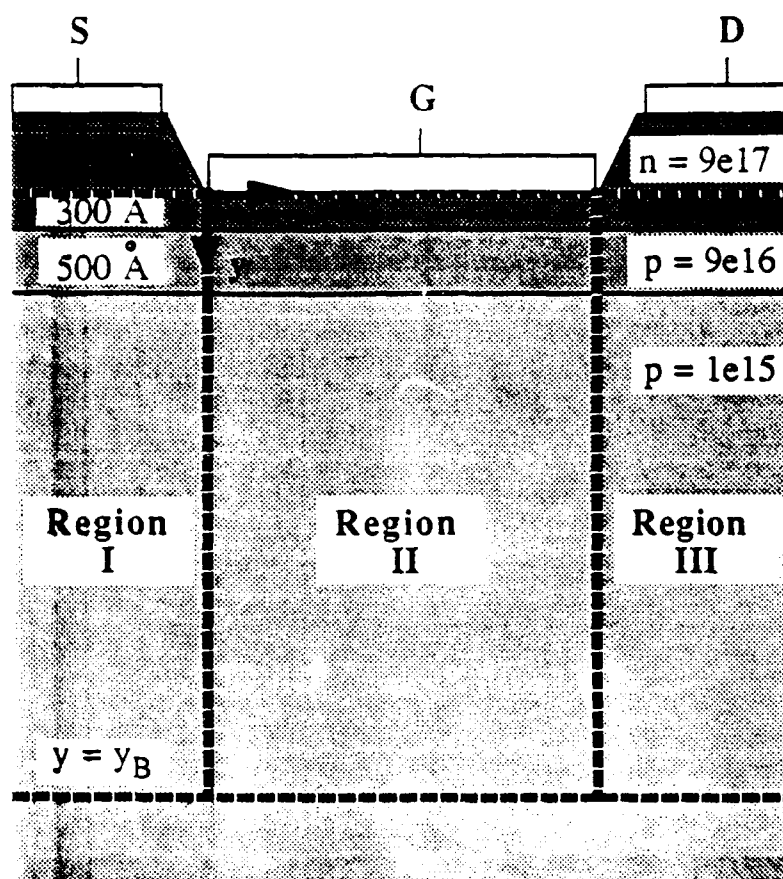


Figure 3.2 Recess-etched MESFET showing the three regions of solution for the 2D Laplace equation.

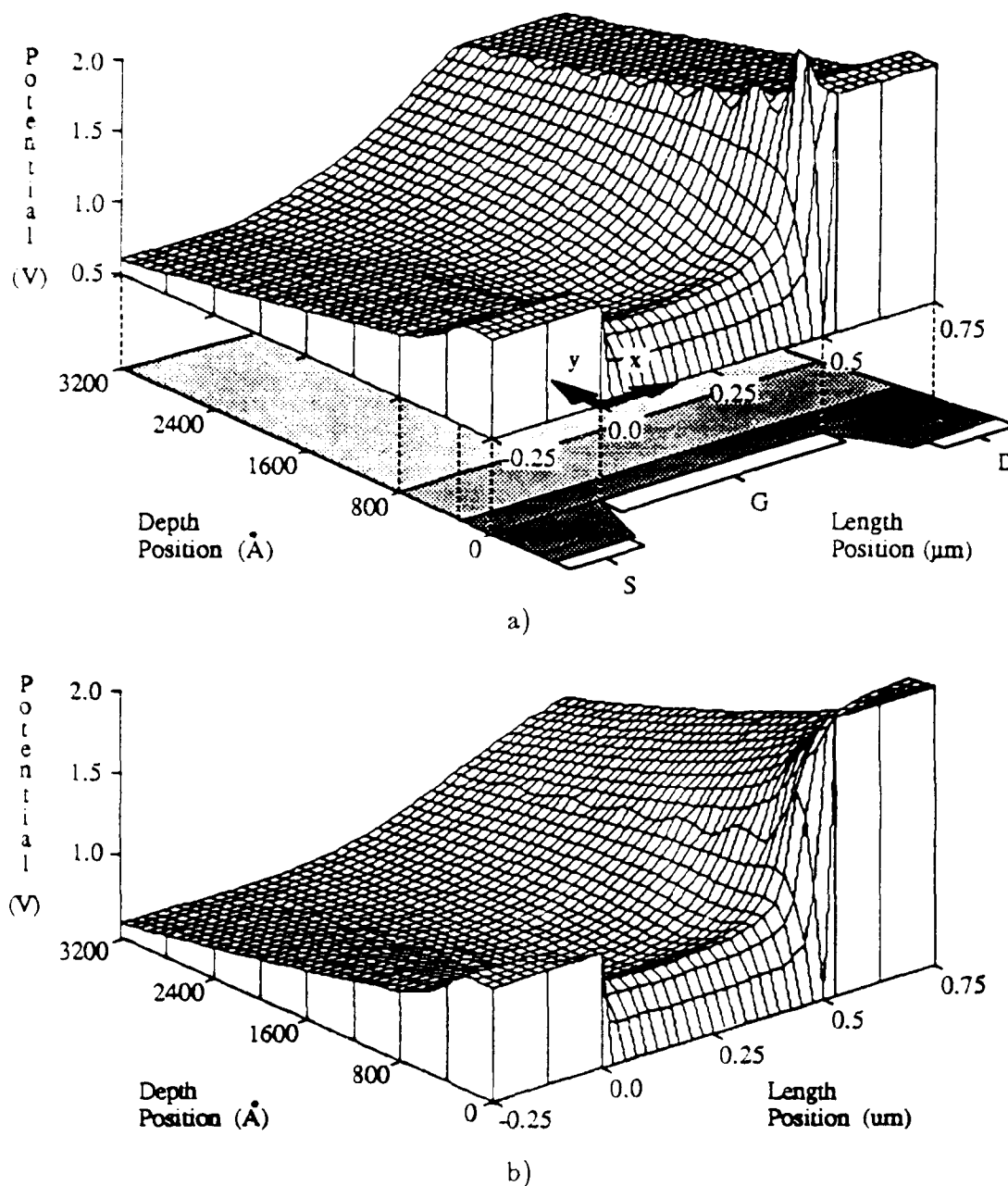


Figure 3.3 Calculated potential in a subthreshold MESFET using 50 harmonics and $\phi = 0$ in the neutral bulk, with $V_{DS} = 1V$.

- a) Harmonic solution only under gate, using 1D Poisson solutions as boundary conditions at source and drain.
- b) Harmonic solutions in source, drain, and gate regions with potential matched at boundaries.

An appropriate choice of y_B depends on the channel length compared to the depletion widths. Consider Figure 3.4a which shows a long-channel device for which the source and drain depletion widths do not overlap. $y_{0,\text{gate}}^{1D}$, $y_{0,\text{source}}^{1D}$, and $y_{0,\text{drain}}^{1D}$ are the depths at which the one-dimensional Poisson equations through the gate, source, and drain predict that the potential will go to zero. For such a long-channel device, it is reasonable to set $\phi_{BB} = 0$ (the potential of the undepleted bulk) and let $y_B = y_{0,\text{gate}}^{1D}$. This condition ensures that the solution area will lie entirely within the depletion region, and it matches the real solution near the middle of the gate where the critical source-to-drain barrier will occur.

Now consider a short-channel device on a lightly doped substrate in which the drain depletion region reaches over to the source as indicated in Figure 3.4b. In such a device, the potential goes to zero nowhere along $y = y_{0,\text{gate}}^{1D}$. A better approximation is achieved by requiring $\phi_{BB} = 0$ at $y_B = y_{0,\text{source}}^{1D}$. The area of the solution still remains entirely within the depletion region, but the position of the back boundary is no longer gate-bias dependent, and the error introduced by the back boundary is moved further from the channel. In order to match Fourier coefficients along the side boundaries between the three regions, the back boundary position chosen for the gate region must be imposed in the source and drain regions also, as is indicated in Figure 3.2.

In the gate region the potential at the metal-semiconductor interface, ($y = 0$), is fixed by the applied gate voltage and the Schottky barrier height. As with the back boundary, the top boundary must be at the same position ($y = 0$) in the source and drain region as it is in the gate region. Because the charge density in the N-layer is treated as changing abruptly at the ends of the gate as previously described, the potential along the top boundary changes abruptly from the gate potential to the potential of the undepleted source or drain (Figure 3.3). This approximation is physically equivalent to having the source and drain ohmic contacts on the same level and immediately adjacent to the gate.

Equation 3.5 is solved in each of the three regions by separation of variables:

Within the gate region: ($0 \leq x \leq L_G$)

$$\phi_G(x,y) = U_{G(0)} + \sum_{n=1}^{\infty} \frac{\sin(nky)}{\sinh(nkL_G)} (C_n^S \sinh(nk(L_G - x)) + C_n^D \sinh(nkx)) \quad [3.8]$$

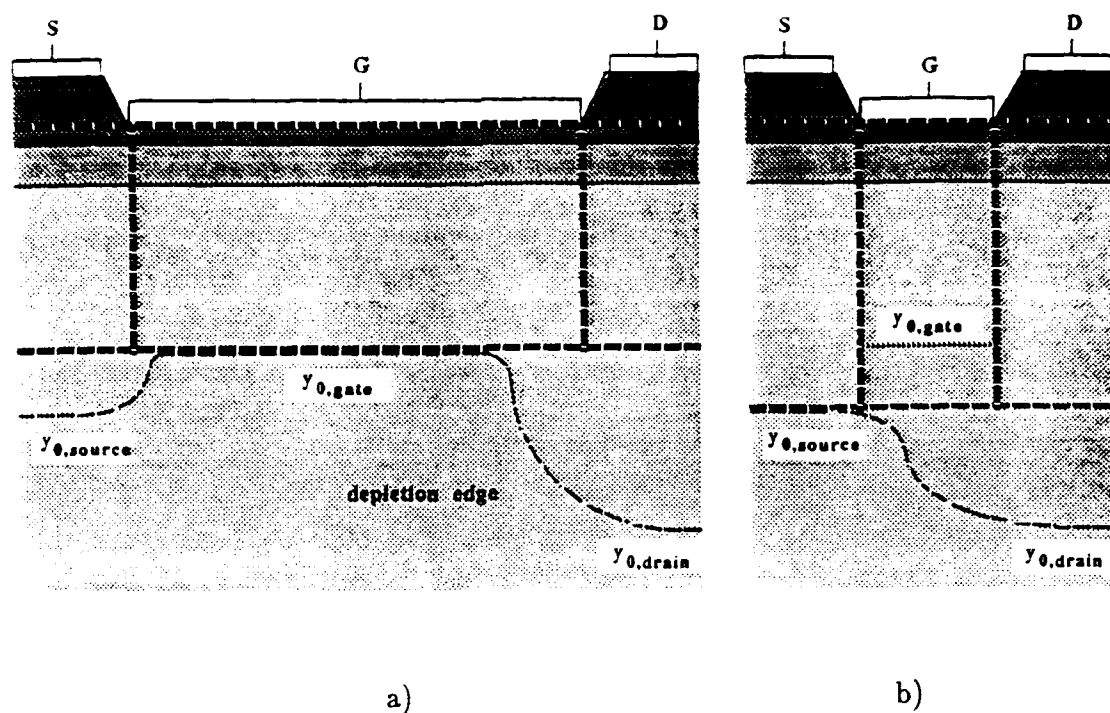


Figure 3.4 Solution regions appropriate for long-channel and short-channel MESFETs.

- a) Long-channel device for which $y_B = y_{0, gate}$.
- b) Short-channel device for which $y_B = y_{0, source}$.

$$\text{where} \quad \frac{d^2 U_G(y)}{dy^2} = -\frac{\rho_G(y)}{\epsilon}, \quad [3.9]$$

$$C_n^S = \frac{2}{y_B} \int_0^{y_B} (\phi(0, y) - U_G(y)) \sin(nky) dy, \quad [3.10]$$

$$C_n^D = \frac{2}{y_B} \int_0^{y_B} (\phi(L_G, y) - U_G(y)) \sin(nky) dy, \quad [3.11]$$

$$\text{and} \quad k = \frac{\pi}{y_B}. \quad [3.12]$$

Within the source region ($x < 0$):

$$\phi_S(x, y) = U_S(y) + \sum_{n=1}^{\infty} \sin(nky) C_n^{S'} \exp(nkx) \quad [3.13]$$

$$\text{where} \quad \frac{d^2 U_S(y)}{dy^2} = -\frac{\rho_S(y)}{\epsilon}, \quad [3.14]$$

$$C_n^{S'} = \frac{2}{y_B} \int_0^{y_B} (\phi(0, y) - U_S(y)) \sin(nky) dy. \quad [3.15]$$

Within the drain region ($x > L_G$):

$$\phi_D(x, y) = U_D(y) + \sum_{n=1}^{\infty} \sin(nky) C_n^{D'} \exp(nk(L_G - x)) \quad [3.16]$$

$$\text{where} \quad \frac{d^2 U_D(y)}{dy^2} = -\frac{\rho_D(y)}{\epsilon}, \quad [3.17]$$

$$C_n^{D'} = \frac{2}{y_B} \int_0^{y_B} (\phi(L_G, y) - U_D(y)) \sin(nky) dy. \quad [3.18]$$

Requiring continuity of ϕ at the boundary between gate-and-source regions ($x = 0$) and at the boundary between the gate-and-drain regions ($x = L_G$) yields:

$$(x = 0): \quad U_G(y) - U_S(y) = \sum_{n=1}^{\infty} (C_n^{S'} - C_n^S) \sin(nky) \quad [3.19]$$

$$(x = L_G): \quad U_G(y) - U_D(y) = \sum_{n=1}^{\infty} (C_n^{D'} - C_n^D) \sin(nky) \quad [3.20]$$

Now define D_n as the difference between the n^{th} coefficients of the solutions on

either side of a boundary:

$$D_n^S = C_n^{S'} - C_n^S = \frac{2}{y_B} \int_0^{y_{11}} (U_S(y) - U_G(y)) \sin(nky) dy \quad [3.21]$$

$$D_n^D = C_n^{D'} - C_n^D = \frac{2}{y_B} \int_0^{y_{11}} (U_D(y) - U_G(y)) \sin(nky) dy. \quad [3.22]$$

The integrals in equations 3.21 and 3.22 involve only known quantities, so the coefficients D_n^S and D_n^D can be calculated directly for any specified top and bottom boundary conditions.

Next the derivative of the potential with respect to x is obtained from equations 3.8, 3.13, and 3.16:

$$\frac{d\phi_G}{dx} = \sum_{n=1}^{\infty} \frac{\sin(nky)}{\sinh(nkL_G)} (nk) (-C_n^S \cosh(nk(L_G - x)) + C_n^D \cosh(nkx)) \quad [3.23]$$

$$\frac{d\phi_S}{dx} = \sum_{n=1}^{\infty} \sin(nky) (nk) C_n^{S'} \exp(nkx) \quad [3.24]$$

$$\frac{d\phi_D}{dx} = \sum_{n=1}^{\infty} \sin(nky) (-nk) C_n^{D'} \exp(nk(L_G - x)) \quad [3.25]$$

Requiring the derivative to be continuous at $x = 0$ and $x = L_G$ yields:

$$C_n^{S'} = -C_n^S \frac{\cosh(nkL_G)}{\sinh(nkL_G)} + C_n^D \frac{1}{\sinh(nkL_G)} \quad [3.26]$$

$$C_n^{D'} = C_n^S \frac{1}{\sinh(nkL_G)} - C_n^D \frac{\cosh(nkL_G)}{\sinh(nkL_G)} \quad [3.27]$$

Finally, solving equations 3.21, 3.22, 3.26, and 3.27 simultaneously gives:

$$C_n^S = \frac{1}{2} D_n^S + K D_n^D \quad [3.28]$$

$$C_n^D = K D_n^S + \frac{1}{2} D_n^D \quad [3.29]$$

$$C_n^{S'} = -\frac{1}{2} D_n^S + K D_n^D \quad [3.30]$$

$$C_n^{D'} = K D_n^S - \frac{1}{2} D_n^D \quad [3.31]$$

$$\text{where } K = \frac{\sinh(nkL_G)}{\exp(2nk) - 1} \quad [3.32]$$

Once the coefficients D_n are calculated from the one-dimensional Poisson solutions $U_G(y)$, $U_S(y)$, and $U_D(y)$ through 3.21 and 3.22, equations 3.28 through 3.32 uniquely specify the Fourier coefficients C_n^S , C_n^D , $C_n^{S'}$, and $C_n^{D'}$. The potential at any point in the gate, source, or drain regions can then be calculated from 3.8, 3.13, or 3.16, respectively. For uniform dopings, as assumed in Figure 3.2, the 1D Poisson equations in each region can be solved analytically, subject to the appropriate contact-imposed boundary conditions at $y = 0$, and to $U = 0$ at $y = y_b$. For non-uniform doping profiles, U_s , U_g , and U_d can be calculated numerically. The electric field at any point can also be obtained efficiently from equations 3.23, 3.24, 3.25, the corresponding summations for $\frac{d\phi}{dy}$, and $\frac{dU}{dy}$.

Figure 3.2 shows the regions of solution for applying the technique to a recess-etched MESFET. Epitaxially-grown JFETs can also be modeled approximately by changing only the gate boundary condition. The Schottky barrier height of the MESFET gate is replaced by the built-in potential of the JFET's PN-junction gate. Technically, the one-dimensional Poisson solution through the gate, $U_G(y)$, should be modified to include the small voltage drop in the P-type gate. This small correction can be ignored for the one-sided P⁺N junctions which commonly form the gates of epitaxially-grown JFETs.

3.2.2 Parameter Extraction from 2-D Potential Solutions

In an N-channel FET biased in the subthreshold region, majority-carrier electrons move from source to drain by thermionic emission and diffusion. Transport across Schottky barriers occurs by the same mechanisms, and theoretical expressions describing current across such one-dimensional potential barriers have been derived for thermionic emission [51], diffusion [52], and combined thermionic emission-diffusion [53]. The theoretical expressions for current density have different pre-factors, but all depend exponentially on the barrier height:

$$J = J_0 e^{-q\phi_b/kT} \quad [3.33]$$

In the two-dimensional problem presented by the subthreshold FET, electrons can move from source to drain along many paths, and each path represents a different barrier height. However, because of the exponential weighting of the barrier height, the total current can be expected to vary directly with the single minimum barrier height to a good approximation. Calculation of the pre-factor (J_0 in 3.33, I_{D50} in 3.2) requires properly weighted summation of the thermionic emission-diffusion current contributions along all possible paths from source to

drain. This is an enormously complicated problem. Fortunately, the insight necessary to design access transistors for minimum subthreshold leakage can be gained by examining the effect of the design parameters on the normalized current density ratio:

$$\frac{J_{DS}}{J_{DS0}} = e^{\frac{-q\phi_{B,MIN}}{kT}} \quad [3.34]$$

Equation 3.34 allows calculation of the normalized current density from the minimum potential barrier height between source and drain, which can be determined for any set of design parameters and operating conditions from the two-dimensional potential solution.

Under most conditions, the minimum potential barrier between source and drain will occur along a path within the N-doped channel region. For example, $\phi_{B,MIN}$ is seen in Figure 3.3b to be the difference between potentials of the undepleted source and the saddle point in the potential ridge along the N-doped channel. Figure 3.5 shows a calculation of normalized current density versus gate voltage using two-dimensional harmonic potential solutions and equation 3.34. The calculation is done for a MESFET with a half-micron gate length, gate barrier height of 0.7 eV, channel doping of 10^{17} cm^{-3} and a channel depth of 1300 Å. Drain-to-source voltage is used as a parameter, and the temperature is 300 K. Comparing Figure 3.5 and Figure 3.1 shows that the two-dimensional calculation produces the expected form of the gate-voltage dependence of the drain-to-source component of the subthreshold FET leakage current.

3.2.2.1 Calculation of Subthreshold Ideality Factor and Subthreshold Swing

Physically, the ideality factor n_s defined in equation 3.2 is the ratio of the change in the minimum barrier height to the change in the gate potential at a fixed drain-to-source voltage. This definition allows n_s to be calculated directly from two-dimensional potential solutions. The subthreshold swing, S , is a related parameter more directly connected to current-versus-gate-voltage plots like Figure 3.5. The subthreshold swing is defined as the change in the gate voltage necessary to change the drain current density by an order of magnitude. The subthreshold swing is a critical parameter because it determines how far below threshold the gate must be taken in order to reduce the drain current to a given value. The subthreshold swing is related to the ideality factor by:

$$S = \frac{kT}{q} \ln(10) n_s \quad [3.35]$$

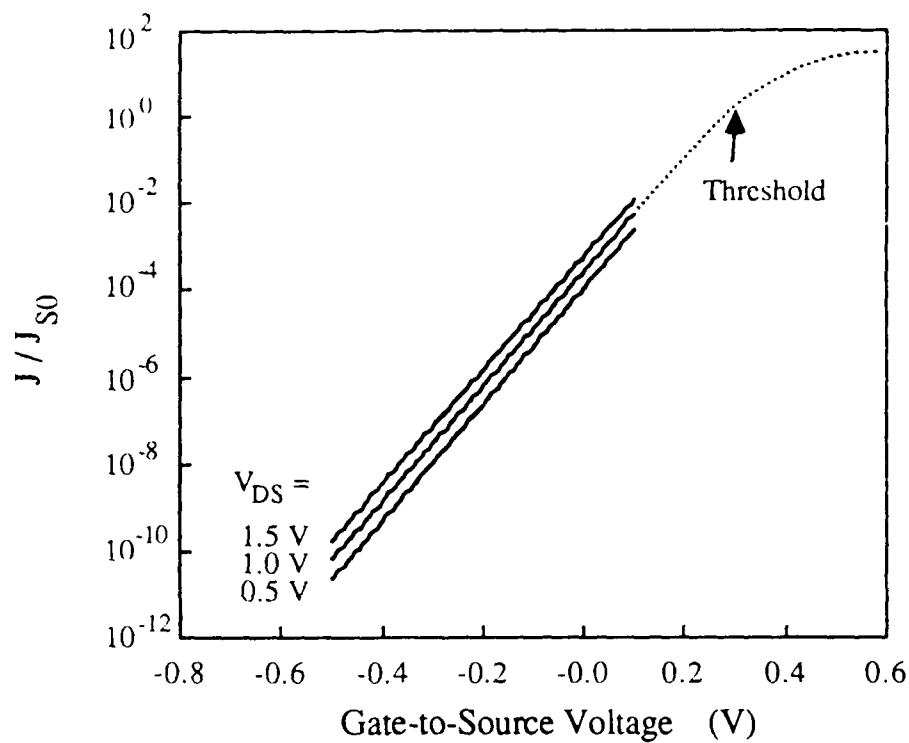


Figure 3.5 Normalized drain current density versus gate voltage, calculated from two-dimensional potential solution. Parameters: $l_G = 0.5 \mu\text{m}$, $N_{DC} = 10^{17} \text{ cm}^{-3}$, $d = 1300 \text{ \AA}$, $\phi_{GC} = 0.7 \text{ V}$, $N_{AB} = 10^{17} \text{ cm}^{-3}$, $w_{PB} = 500 \text{ \AA}$, $N_{AI} = 10^{15} \text{ cm}^{-3}$, $T = 300 \text{ K}$. (Parameters are defined in Figure 3.7.)

The minimum ideality factor occurs in an ideal device where any decrease in gate potential directly increases the drain-to-source barrier so that $n_s = 1$. Equation 3.35 then shows that the minimum subthreshold swing for an ideal device is $kT/q \ln(10)$, which equals about 60 mV/decade at room temperature. As temperature increases, S increases, indicating that the drain leakage current at a certain gate voltage at high temperatures will exceed the leakage at the same gate voltage at low temperature.

3.2.2.2 Calculation of Threshold Voltage

To extract the threshold voltage from two-dimensional potential calculations, the threshold voltage is defined as the gate voltage which will reduce the minimum potential barrier height to some specified value. The specific value chosen is somewhat arbitrary, corresponding to arbitrarily specifying a small drain current (e.g., 1 μA) to signify the threshold condition in a FET. In equation form, the threshold voltage is defined by:

$$\phi_{B,\text{MIN}}(V_G=V_T) = c \frac{kT}{q} - (E_C - E_F) \big|_{\text{source}} \quad [3.36]$$

where c is an arbitrary constant. In all calculations in this work, c is chosen to be 4.0. To calculate the threshold voltage for a specified device at a given temperature and drain-to-source voltage, two-dimensional potential solutions are iteratively performed for different gate voltages until equation 3.36 is satisfied.

The parameter V_{T0} is defined in equation 3.3 as the threshold voltage extrapolated to zero drain-to-source bias. Because the threshold voltage is defined in terms of potentials instead of currents in equation 3.36, V_{T0} can be calculated directly without extrapolating by setting V_{DS} to zero and iterating. The parameter κ of equation 3.3 can be determined from the shift in calculated threshold voltages caused by applying different drain-to-source biases. Figure 3.6 shows a calculation of threshold voltage versus drain-to-source bias for the same transistor parameters used in Figure 3.5. The linear dependence specified by equation 3.3 is apparent. The values of V_{T0} and κ calculated for this device are 0.336 V and -0.044, respectively.

3.2.3 Calculated Dependences of Subthreshold Parameters on Dimensions and Dopings

This section will present calculations of the subthreshold parameters (V_{T0} , κ , n_s , and S) as functions of the physical design parameters. The objective is to illustrate the effects of varying the design parameters, and also to demonstrate the capabilities of the harmonic two-dimensional potential solution technique.

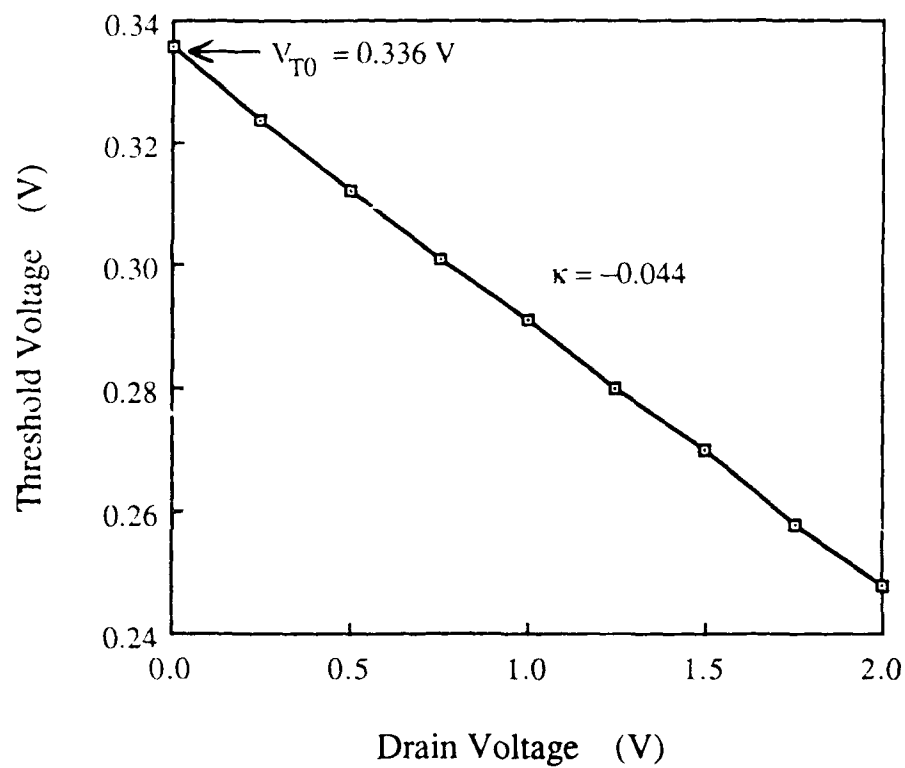


Figure 3.6 Threshold voltage versus drain-to-source voltage, calculated using the parameters of Figure 3.5.

The physical parameters which may be selected in the design of a JFET or MESFET access transistor are the dopings and thicknesses of the epitaxially-grown layers and the horizontal dimensions such as the channel length and width. Figure 3.7 defines the important parameters. The P-type layer lying just below the N-type channel is designed to be completely depleted to maintain low parasitic capacitances. Its purpose is to increase the potential barrier which defines the channel, suppressing substrate leakage and short-channel effects. The layer underneath is assumed to be a nominally-undoped epitaxial layer with a background net acceptor concentration of approximately 10^{15} cm^{-3} .

The threshold voltage of a JFET or MESFET will be determined by the channel doping (N_{DC}) and depth (d), once the gate-to-channel barrier height (ϕ_{GC}) is known. There is also a dependence on the thickness (w_{PB}) and doping (N_{AB}) of the depleted P layer, although the depletion of the channel from the lower junction is typically small. The threshold voltage also depends on V_{DS} as discussed in Section 3.2.2.2 and shown in Figure 3.6. Finally, even the zero- V_{DS} threshold voltage, V_{T0} , depends on channel length for very short-channel transistors. Figures 3.8 and 3.9 show calculations of MESFET and JFET long-channel ($2 \mu\text{m}$) zero- V_{DS} threshold voltages versus channel depth and doping. For both calculations, $N_{AB} = 10^{17} \text{ cm}^{-3}$ and $w_{PB} = 500 \text{ \AA}$. For the MESFET, $\phi_{GC} = 0.7 \text{ V}$, and the gate doping of the JFET is 10^{19} cm^{-3} .

The variation of the zero- V_{DS} threshold voltage with length for very short-channel devices is shown in Figures 3.10 and 3.11. Both calculations are for MESFETs with gate-to-channel barrier heights of 0.7 V . Figure 3.10 represents enhancement mode devices, while Figure 3.11 depicts depletion mode devices. Each figure includes calculations for channel dopings of 1 , 5 , and $9 \times 10^{17} \text{ cm}^{-3}$, with the channel depths adjusted so that the long-channel V_{T0} values are $+0.3 \text{ V}$ and -1.0 V for Figures 3.10 and 3.11 respectively. In each figure, it is apparent that increasing the channel doping reduces the zero- V_{DS} threshold shift at very short gate lengths. The smaller sensitivity of the enhancement mode devices to short channel effect will be discussed in Section 3.2.4.

The parameter κ describing the effect of drain-to-source voltage on threshold voltage is also dependent on gate length and channel doping. Figures 3.12 and 3.13 show calculations of κ versus gate length for the same sets of devices depicted in Figures 3.10 and 3.11. Again, heavy channel doping is seen to suppress the sensitivity of the threshold voltage to the drain-to-source bias. Values of κ with magnitude greater than about 0.1 result in transistors which are difficult to use in digital circuits, so Figures 3.12 and 3.13 illustrate that it is necessary to use heavy channel dopings to achieve very short gate lengths, particularly for depletion mode devices.

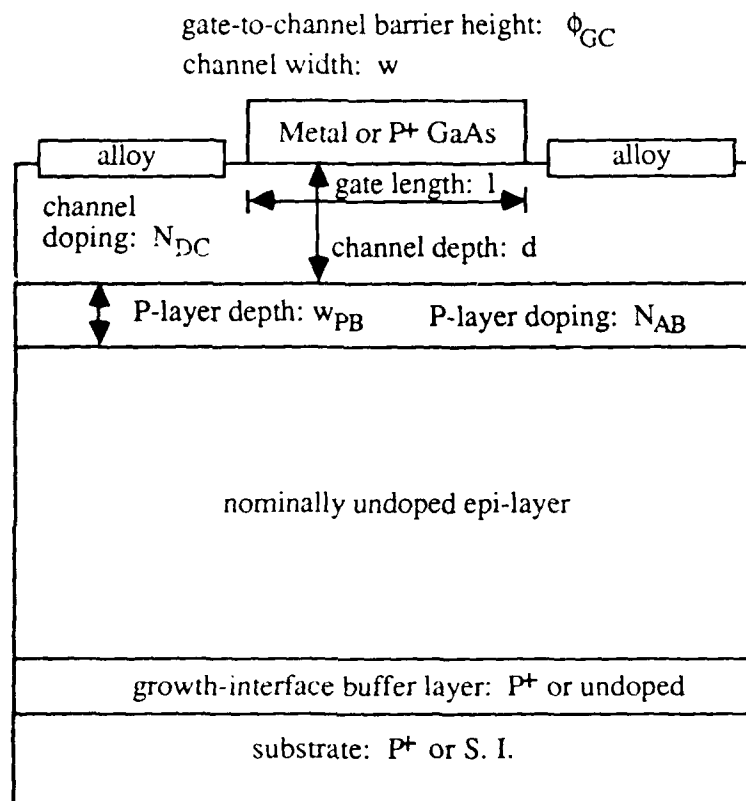


Figure 3.7 Definition of physical design parameters for JFET and MESFET access transistors.

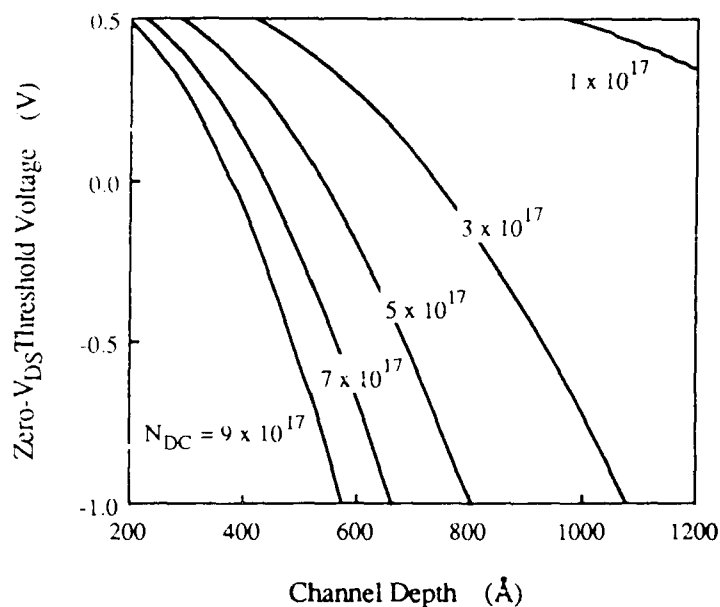


Figure 3.8 MESFET V_{T0} versus channel depth with channel doping as a parameter. $\phi_{GC} = 0.7$ V, $l_G = 2.0$ μm , $w_{PB} = 500$ \AA , $N_{AB} = 10^{17}$ cm^{-3} , $N_{AI} = 10^{15}$ cm^{-3} , $T = 300$ K.

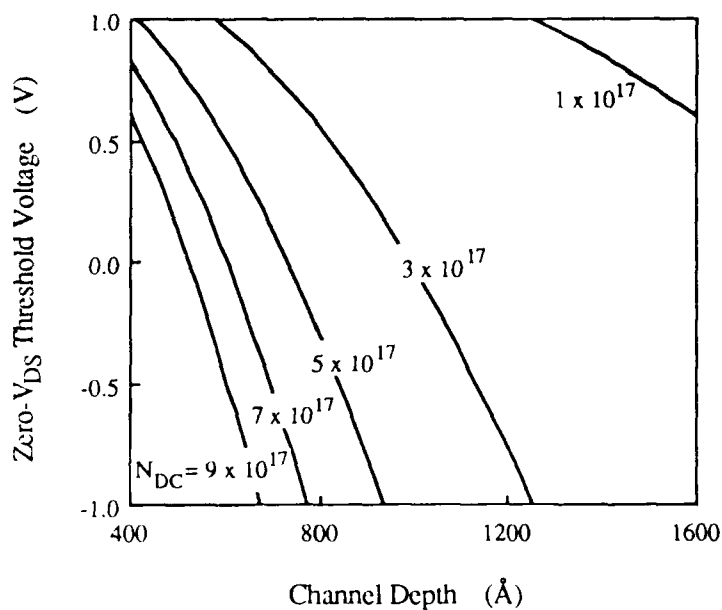


Figure 3.9 JFET V_{T0} versus channel depth with channel doping as a parameter. $\phi_{GC} = 1.4$ V, $l_G = 2.0$ μm , $w_{PB} = 500$ \AA , $N_{AB} = 10^{17}$ cm^{-3} , $N_{AI} = 10^{15}$ cm^{-3} , $T = 300$ K.

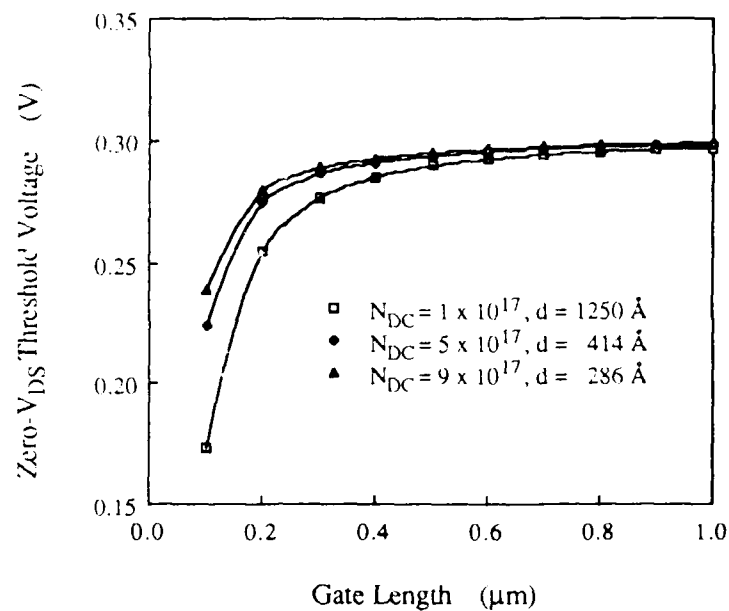


Figure 3.10 V_{T0} versus gate length for an enhancement-mode MESFET. Parameters: $\phi_{GC} = 0.7 \text{ V}$, $N_{AB} = 10^{17} \text{ cm}^{-3}$, $w_{PB} = 500 \text{ \AA}$, $N_{AI} = 10^{15} \text{ cm}^{-3}$, $T = 300 \text{ K}$, channel doping and depth as indicated.

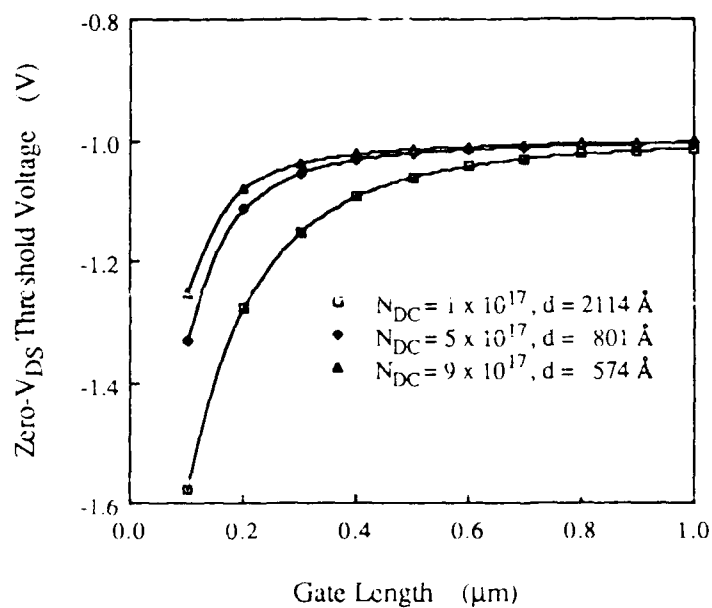


Figure 3.11 V_{T0} versus gate length for a depletion-mode MESFET. Parameters as in Figure 3.10.

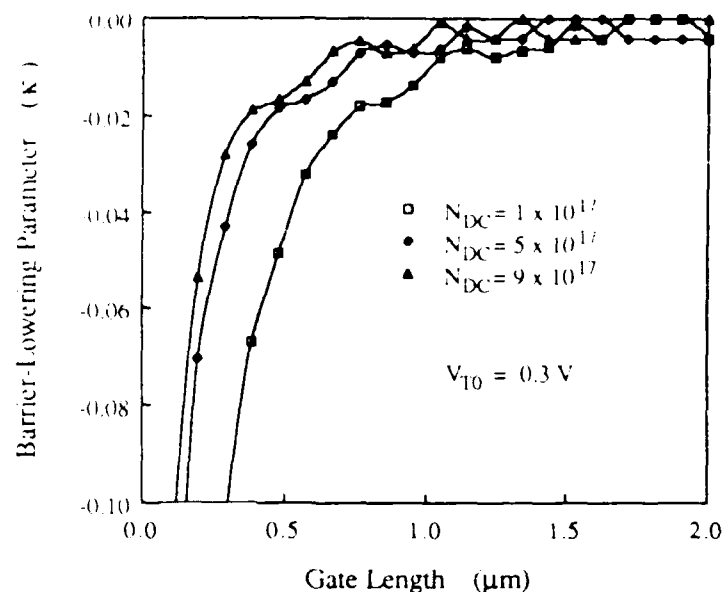


Figure 3.12 κ versus gate length for an enhancement-mode MESFET. Channel doping as indicated. Channel depth adjusted to maintain $V_{T0} = 0.3$ V. Other parameters as in Figure 3.10.

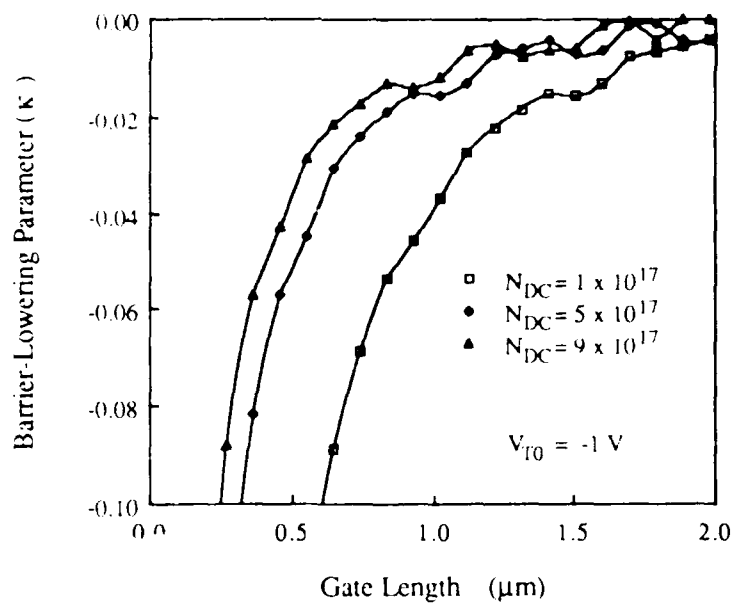


Figure 3.13 κ versus gate length for a depletion-mode MESFET. Channel depth adjusted to maintain $V_{T0} = -1.0$ V. Other parameters as in Figure 3.10.

Figure 3.14 shows a calculation of the effects of channel doping on the subthreshold ideality factor, n_s . The two curves in the figure represent enhancement and depletion mode MESFETs with $0.5\text{ }\mu\text{m}$ gate lengths. The upper curve is for depletion devices with $V_{T0} = -1.0\text{ V}$, and the lower curve is for enhancement devices with $V_{T0} = 0.3\text{ V}$. The channel depths were adjusted at each doping to maintain the same threshold voltage. Increased channel doping reduces the subthreshold ideality factor, corresponding to increased slope on a $\log(I_D)$ versus V_{GS} plot. The equivalent values of the subthreshold swing, S , at 300 K are shown on the second vertical axis. The advantage of the enhancement-mode device over the depletion-mode device will be discussed in Section 3.2.4.

The subthreshold ideality factor also increases with reduced gate length. Figure 3.15 plots ideality factor versus gate length for enhancement and depletion mode MESFETs. For both curves, the channel doping is $5 \times 10^{17}\text{ cm}^{-3}$. As in Figure 3.14, the channel depth was adjusted to maintain threshold voltages of 0.3 V and -1.0 V for the enhancement and depletion devices, respectively.

The doping of the layers underlying the channel also affects the subthreshold characteristics. Intentionally doping the lowest epilayer significantly reduces short channel effects, but it greatly increases parasitic capacitances. Increased capacitances slow the circuit operation, negating the beneficial effects of shortened gate lengths. It will be assumed that the lowest epilayer is not intentionally doped, so that its doping is not a design parameter.

The doping and thickness of the depleted P layer just beneath the channel may be selected, but the total amount of charge must remain sufficiently small so that the total depletion depth is not significantly reduced. The effects of this layer are not easily quantified from the two-dimensional potential solutions. This is because the major function of this layer is not to increase the subthreshold barrier height, but to confine the current to the channel. Figure 3.16 illustrates this function with calculated conduction band edges along the y direction of Figure 3.2 for MESFETs with and without the 500 \AA , 10^{17} cm^{-3} -doped P layer. Although the channel minima occur at the same energy for both devices, the minimum is much broader for the MESFET without the depleted P layer. The current flowing in the undoped material under the channel would be considerably greater in the device without the P layer. By confining the current to the N-doped channel, where it is most affected by the gate bias, the depleted P layer improves the subthreshold characteristics. However, because it is a current confinement effect not easily calculated by the harmonic solution technique, the depth and doping of the P layer will not be considered as design

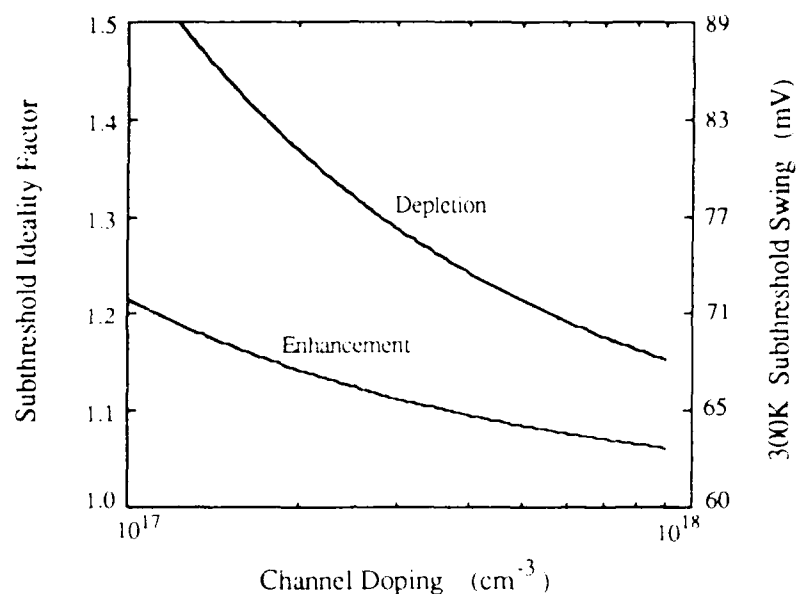


Figure 3.14 Subthreshold ideality factor versus channel doping for enhancement and depletion MESFETs. Channel depths adjusted to maintain $V_{T0} = 0.3 \text{ V}$ and -1.0 V . Other parameters as in Figure 3.10.

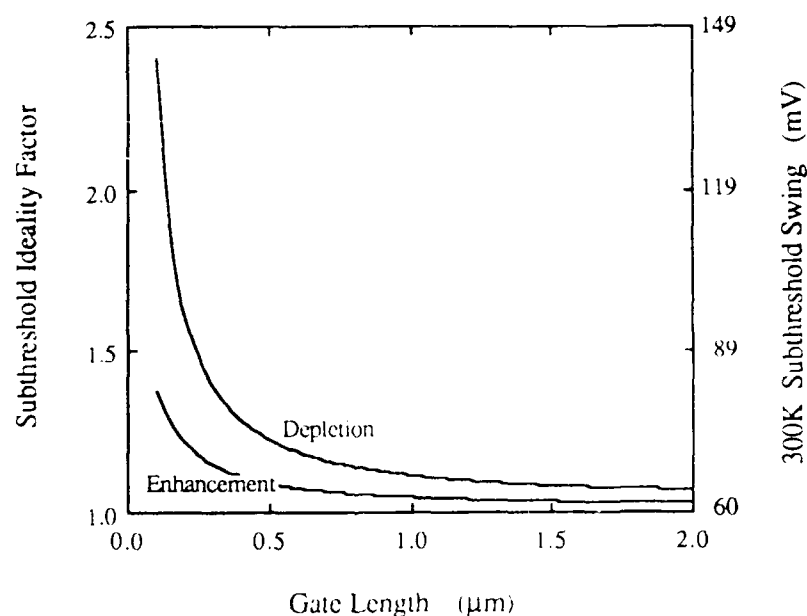


Figure 3.15 Subthreshold ideality factor versus gate length for enhancement and depletion MESFETs. Channel doping $= 5 \cdot 10^{17} \text{ cm}^{-3}$. Channel depths adjusted to maintain $V_{T0} = 0.3 \text{ V}$ and -1.0 V . Other parameters as in Figure 3.10.

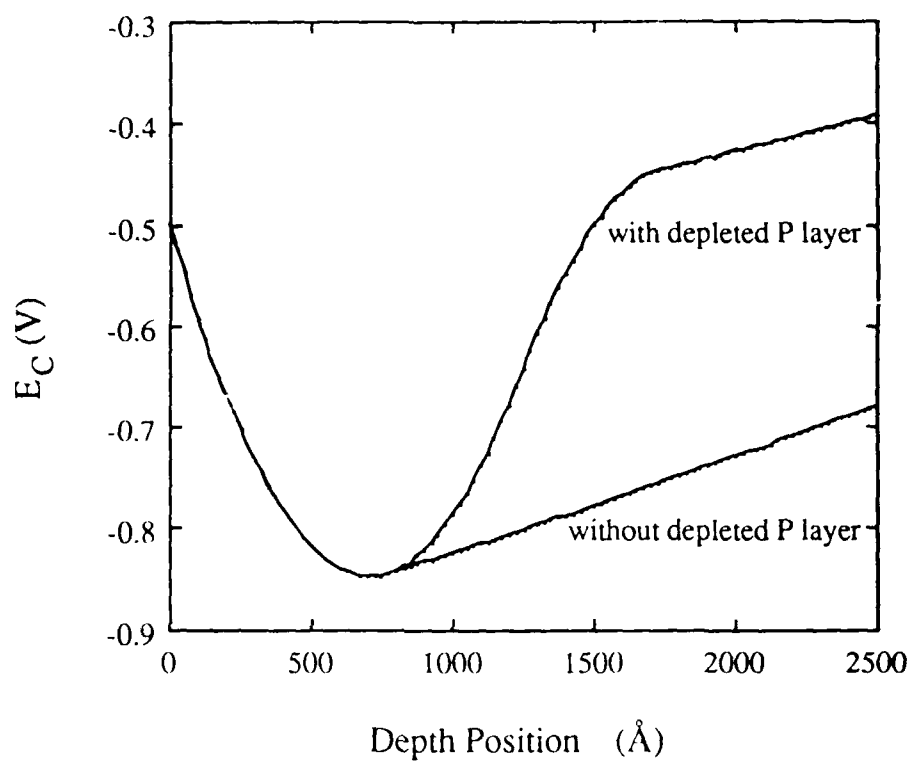


Figure 3.16 Conduction band cross sections through the gates of MESFETs with and without the depleted P layer underlying the channel. E_C equals zero in the grounded substrate. Parameters: $N_{DC} = 10^{17} \text{ cm}^{-3}$, $d = 1238 \text{ Å}$ (with P layer), 768 Å (without P layer), $\phi_{GC} = 0.7 \text{ V}$, $N_{AB} = 10^{17} \text{ cm}^{-3}$, $w_{PB} = 500 \text{ Å}$, $N_{Al} = 10^{15} \text{ cm}^{-3}$.

parameters. $N_{AB} = 10^{17} \text{ cm}^{-3}$ and $w_{PB} = 500 \text{ \AA}$ have been used in all the calculations presented so far. It is possible to achieve the same benefits in epitaxially-grown FETs by using a heterojunction to confine the channel [54].

3.2.4 Selection of Design Parameters for Optimum Subthreshold Performance

In an ideal device, every incremental decrease in the gate potential would produce an identical increase in the potential barrier separating the source and drain. In real devices with short gate lengths, the fixed source and drain potentials tend to hold the channel potential in place, resisting the effects of changes in gate voltage. This can be combatted in two ways. One way is to increase the doping of any of the layers, particularly the channel, so that the potentials from the source and drain do not encroach as far. A second way is to change the design so that the path with the minimum potential barrier is a shorter distance from the gate. This effectively increases the gate-to-channel capacitance and results in better subthreshold performance.

The sets of curves in Figures 3.12 and 3.13 show both effects occurring as the channel doping is increased. Increasing doping requires reduction of the channel depth to maintain the same value of V_{T0} . As the depth is decreased, the short channel effects become less severe for both the enhancement and depletion devices. Comparing the two figures shows that enhancement-mode devices perform better than depletion devices with the same channel doping. Given enhancement and depletion transistors with the same channel doping and the same source-to-drain barrier, the minimum barrier will occur closer to the gate in the enhancement-mode device. This situation is illustrated by Figure 3.17, which shows conduction band cross-sections through the gates of an enhancement and a depletion MESFET. The transistors are identical except for the channel depth which has been adjusted to produce V_{T0} values of 0.3 V and -1.0 V. Each transistor is biased 0.1 V below threshold, so that the channel minima occur at the same energy. The distance from the channel minimum to the gate in the enhancement device is only half as large as in the depletion-mode device, resulting in improved subthreshold performance.

The same effect causes transistors with lower gate-to-channel barriers to have larger subthreshold slopes than otherwise identical devices with large ϕ_{GC} . Figure 3.18 plots subthreshold ideality factor for a JFET and a MESFET with identical threshold voltages and channel dopings. The smaller subthreshold swings in the MESFET are due to smaller separation between the minimum source-to-drain barrier and the gate.

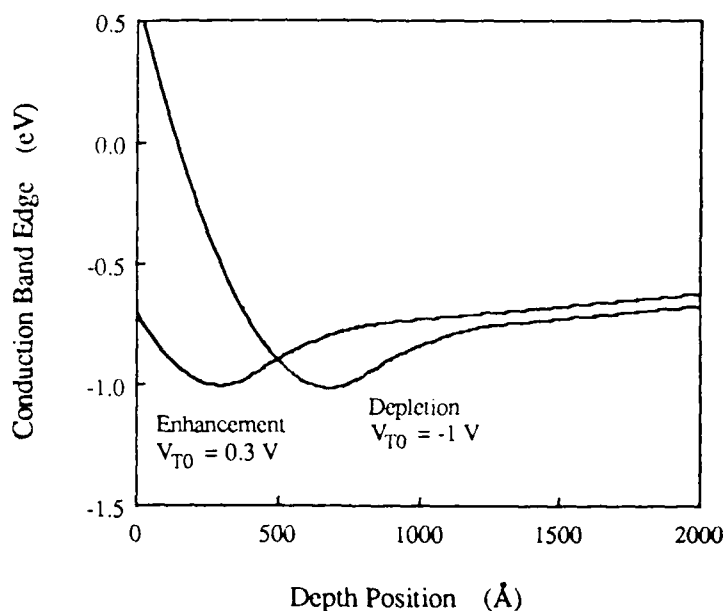


Figure 3.17 Conduction band cross-sections through the gates of enhancement and depletion MESFETs. $V_{T0} = 0.3$ V, -1.0 V. E_C equals zero in the grounded substrate. $V_{GS} = V_T - 0.1$ V for each device. Channel doping = $5 \times 10^{17} \text{ cm}^{-3}$.

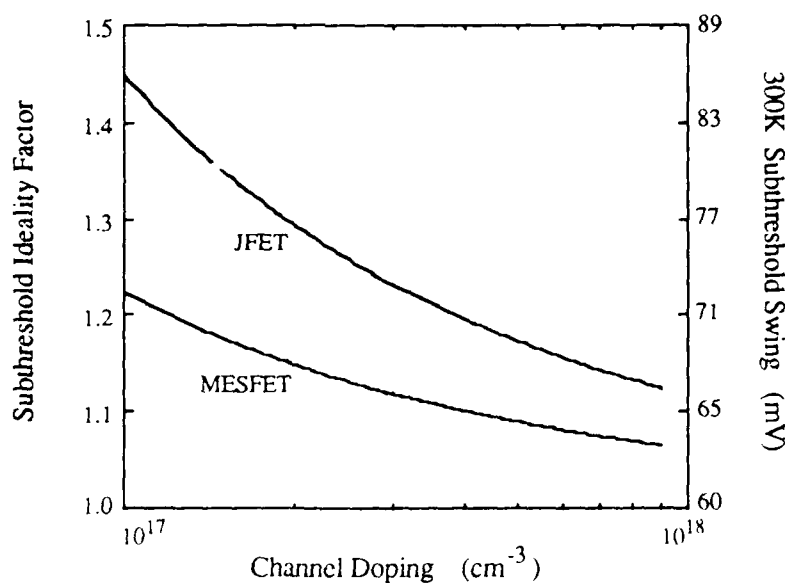


Figure 3.18 Subthreshold ideality factor for a JFET and MESFET. $l_G = 0.5 \mu\text{m}$ and channel depth is adjusted to maintain $V_{T0} = 0.3$ V for each device.

To produce a transistor with a given threshold voltage, increasing the channel doping or lowering the gate-to-channel barrier height will improve the subthreshold characteristics. Increasing channel doping results in decreased carrier mobility. More importantly, increasing channel doping increases the sensitivity of threshold voltage to channel depth, as shown in Figures 3.8 and 3.9. This means that threshold voltage control becomes increasingly difficult, reducing yields for large circuits. There is no question, however, that heavy channel doping must be employed. It is necessary, both to allow short gate lengths for high-speed operation, and to reduce thermal-generation leakage currents as described in the previous chapter. The maximum practical channel doping will be established by the observed trade-off of manufacturability and performance.

Lowering the gate-to-channel barrier height sharply increases the gate-leakage component of the drain current, as discussed in Section 3.2. JFET gate leakage should consist only of the same mechanisms which are present in the PN-junction storage capacitor. The storage capacitors have demonstrated excellent storage time performance, as reported in the previous chapter, so JFET gate-to-channel barrier heights should be more than sufficient. The feasibility of reducing the JFET gate-to-channel barrier height to that of a treated-gate MESFET or of a simple MESFET is currently being evaluated experimentally.

Once channel doping and gate-to-channel barrier heights have been established, the subthreshold parameters are still a function of the threshold voltage selected. Fortunately, enhancement-mode devices exhibit better characteristics, and they are also desirable for the power dissipation advantages discussed in Chapter 1. The obstacles to employing enhancement-mode access transistors lie in finding a cell geometry which can efficiently combine an enhancement-mode access transistor with the required depletion-mode capacitor, and in the reduced noise margins associated with an enhancement-access cell. These considerations will be discussed in Chapter 4.

3.3 Measured FET Subthreshold Performance

This section reports preliminary measurements of the subthreshold performance of epitaxially-grown mesa-isolated JFETs and MESFETs. The data are preliminary in that the processing sequences, dimensions, and epitaxial layers are currently being modified to improve the results. However, the measured characteristics do illustrate some of the theoretical dependences described in the previous section.

Figure 3.19 shows a measured drain current versus gate-to-source voltage characteristic for a $5\text{ }\mu\text{m}$ -gate-length JFET. Figure 3.20 shows the same characteristic for an ammonium-sulfide-treated MESFET of the same dimensions. Both devices were measured at drain-to-source biases of 1.0, 1.5, and 2.0 V. The curves were taken at room temperature in the dark. Both transistors are ring-gate access transistors from experimental DRAM cells, with effective gate lengths of $300\text{ }\mu\text{m}$ and gate areas of approximately $2500\text{ }\mu\text{m}^2$. The ring-gate structure and its purpose will be discussed in Chapter 4.

The JFET of Figure 3.19 is depletion mode, with a threshold voltage of about -0.25 V . The MESFET of Figure 3.20 is enhancement mode with a threshold of about $+0.3\text{ V}$. As described in the previous section, both the lower gate-to-channel barrier and the more positive threshold of the MESFET tend to increase its subthreshold slope. This is observed experimentally, with the JFET showing a subthreshold swing of 75 mV/decade while the subthreshold swing of the MESFET is 63 mV/decade . At room temperature, these subthreshold swings correspond to subthreshold ideality factors of 1.26 and 1.06 for the JFET and the MESFET, respectively.

The beneficial effects of the larger gate-to-channel barrier of the JFET can be seen in its ability to handle gate biases farther above threshold, and in its lower gate-to-drain leakage. The gate-to-drain leakage component in the JFET is not observable in Figure 3.19. The drain current for $V_G < -0.5\text{ V}$ actually consists of drain-to-substrate leakage, rather than leakage from the gate. The gate-to-drain leakage component due to thermal generation is orders of magnitude smaller. The JFET's drain-to-substrate leakage component will be discussed later in this section.

The drain current in the MESFET for $V_G < -0.2\text{ V}$ can be attributed directly to gate-to-drain leakage by thermionic emission. The minimum of 280 pA in the $V_{DS} = 1\text{ V}$ drain current for the MESFET corresponds to a gate current density of about $1.2 \times 10^{-13}\text{ A}/\mu\text{m}^2$. This means that if the leakage rate scales directly with area, then the minimum current for a $1 \times 10\text{ }\mu\text{m}$ gate would be only 1.2 pA at room temperature. The MESFET's gate-to-drain leakage component depends fairly strongly on gate-to-drain bias. If the threshold voltage of the MESFET were translated to -1.0 V , Figure 3.20 shows that the minima in the drain current characteristics would increase by more than an order of magnitude.

As expected for $5\text{ }\mu\text{m}$ gates, neither of the FETs in Figures 3.19 and 3.20 show any short-channel effects. The source-to-drain leakage current is independent of drain-to-source voltage. Figure 3.21 shows measured drain

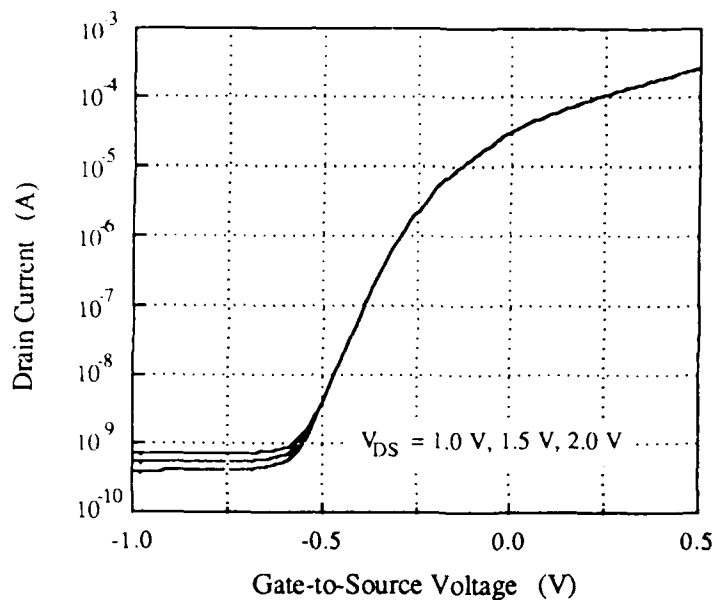


Figure 3.19 Measured drain current versus gate-to-source voltage for an epitaxial JFET with $V_T = -0.25$ V. Gate length = $5 \mu\text{m}$, effective width = $300 \mu\text{m}$, gate area = $2500 \mu\text{m}^2$. Subthreshold swing = 75 mV/decade.

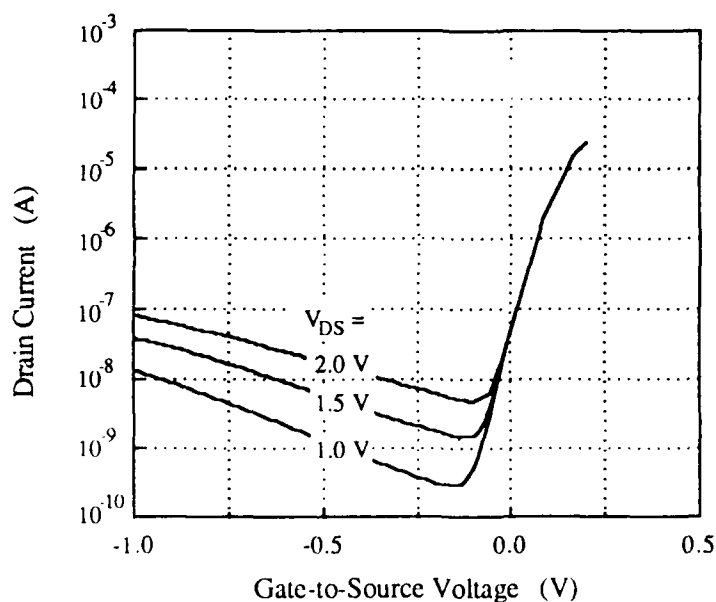


Figure 3.20 Measured drain current versus gate-to-source voltage for an epitaxial $(\text{NH}_4)_2\text{S}$ -treated MESFET with $V_T = +0.3$ V. Gate length = $5 \mu\text{m}$, effective width = $300 \mu\text{m}$, gate area = $2500 \mu\text{m}^2$. Subthreshold swing = 63 mV/decade.

current versus gate-to-source voltage for a $0.25\text{ }\mu\text{m}$ gate JFET in which the short channel effects are readily apparent. The threshold voltage has been increased by more than a volt compared to the long-channel JFET with the same channel depth and doping. The subthreshold swing has increased to 1.92V/decade , corresponding to an ideality factor of 32. The exhibited variation of the threshold voltage with drain-to-source bias implies that $\kappa = -0.48$. the minimum drain current has increased by three orders of magnitude. While this transistor is not optimized and its effective gate length may be less than $0.25\text{ }\mu\text{m}$, its poor subthreshold performance definitely points out the challenges in designing workable sub-micron DRAM access transistors.

Figures 3.22 and 3.23 show the terminal current characteristics of a $2\times 20\text{ }\mu\text{m}$ gate JFET. In the first figure, the substrate is tied to the grounded source, while the drain is at one volt. The drain-to-substrate leakage can be seen in the two order-of-magnitude difference in the drain and source currents for $V_G < -1.5\text{ V}$. With the substrate and source both grounded, the source current consists only of leakage from the gate. The high leakage of the drain-to-substrate diode in these JFETs can be attributed to spiking of the contact alloy through the channel layer, forming a Schottky barrier to the substrate. Future JFET designs hope to avoid this difficulty by using a heavy, shallow N-type implant of the contacts before metal deposition to reduce the need for a high-temperature alloying step.

A more fundamental problem is apparent in Figure 3.22. Below threshold the gate depletion region punches through to the substrate and a large gate-to-substrate current begins to flow. Figure 3.23 shows that the gate punchthrough current can be eliminated by biasing the substrate with the gate. When the substrate is biased with the gate, both the drain and source leak to the substrate and to the gate. The figure shows that the leakage from the spiked alloyed contacts produces a substrate current of $>100\text{ pA}$, while the thermal-generation gate current is down at the detectability limit of 1 pA . The effects of gate-to-substrate punchthrough current on DRAM cell design will be considered in Chapter 4.

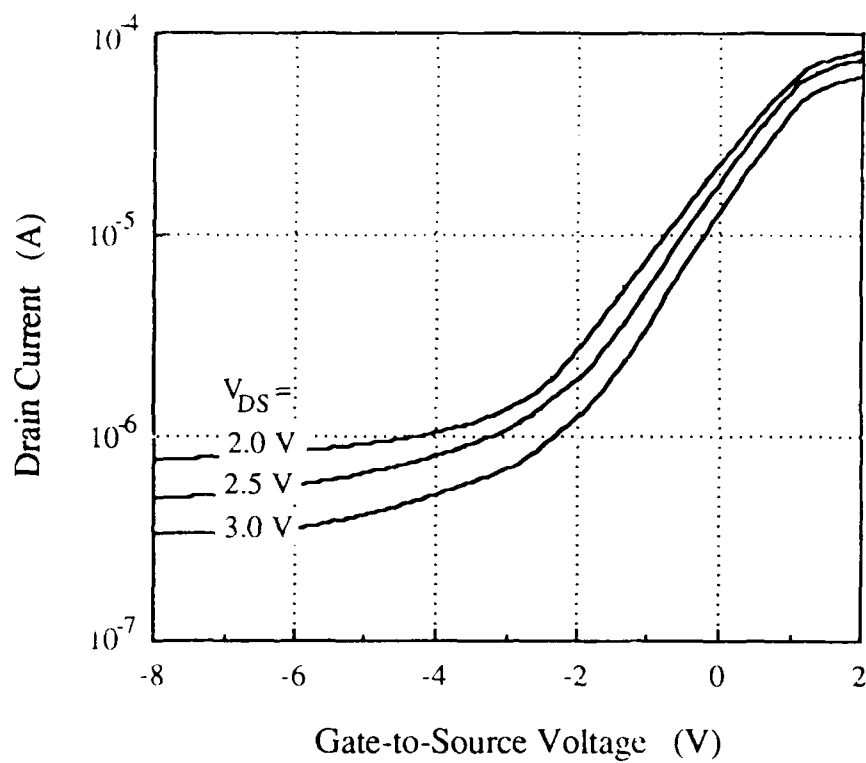


Figure 3.21 Measured drain current versus gate-to-source voltage for a $0.25 \mu\text{m} \times 20 \mu\text{m}$ JFET.

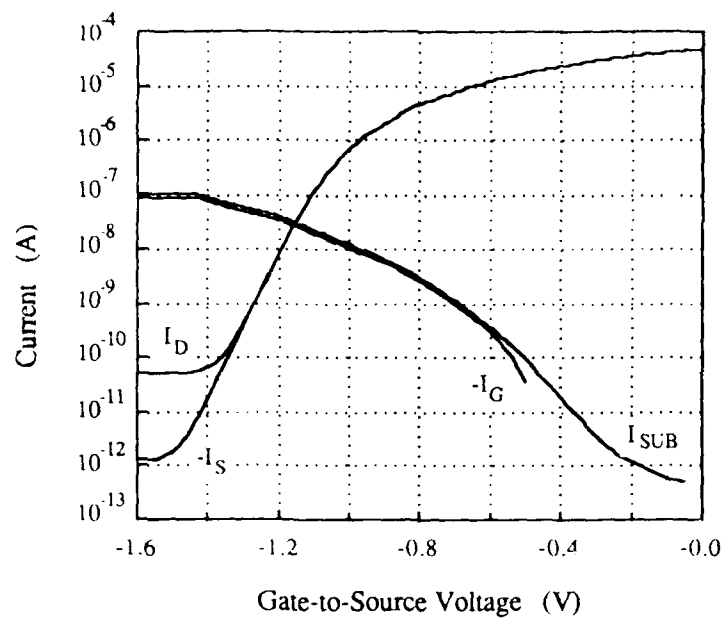


Figure 3.22 Measured terminal currents in a $2 \times 20 \mu\text{m}$ JFET with the substrate grounded, showing gate-to-substrate punchthrough current.

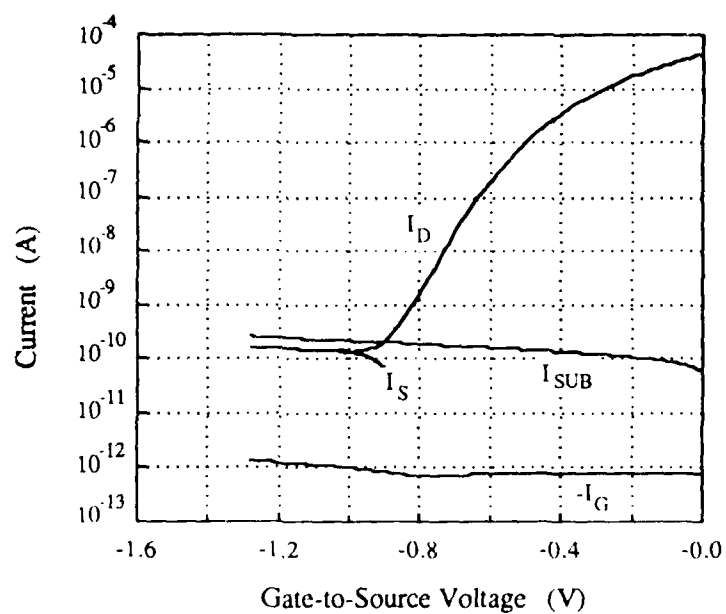


Figure 3.23 Measured terminal currents in a $2 \times 20 \mu\text{m}$ JFET with the substrate biased with the gate.

CHAPTER 4

COMPLETE DYNAMIC MEMORY CELLS

4.1 Operating Voltages

Selection of operating voltages for a GaAs FET-accessed dynamic memory cell is more complicated than in a Si design, because without a gate insulator the voltages must be selected to avoid significant gate currents. The gate diodes of JFET and MESFET access transistors should not be taken too far into forward or reverse bias. During read and write sequences, the forward bias on the gate controls the capacitor-to-bitline current, and thus the cycle time. In the storage state, the reverse bias on the gate must be large enough to reduce the capacitor-to-bitline leakage current to an acceptable value. The objective of this section is to describe wordline, bitline, and logic-level voltage requirements for the read, write, and storage cycles of FET-accessed cells.

The minimum capacitor-to-bitline current necessary during read/write cycles and the maximum current permissible in the storage state will be determined by the amount of signal charge and the desired cycle times. To achieve maximum speed and single-chip capacity, the signal charge may be as small as 100 fC. To compete with the performance of other technologies, the read and write times must be near one nanosecond. To allow a 1 KHz refresh rate, the cell must retain 90% of the stored signal charge for 1 ms over the desired operating temperature range. These figures mean that the access transistor must be capable of supplying capacitor-to-bitline currents ranging from 10 pA to hundreds of μA , a range of more than seven orders of magnitude.

In order to obtain such a large range of currents, the gate voltages applied to the access transistors within the memory array will have to vary over a much larger range than is used in standard GaAs direct-coupled logic. The logic-level voltages used within the memory array will not coincide with the standard logic levels used in the peripheral circuitry. The sense-amplifiers must transform the internal memory array logic levels back into standard values. High density silicon dynamic memories are often constructed similarly, with different internal and peripheral logic levels [55].

In this discussion, V_{on} is defined as the voltage above threshold which must be applied to the gate to achieve the desired bitline-to-capacitor current densities during reading and writing. V_{off} is the magnitude of the voltage below threshold which must be applied to the gate to reduce the bitline-to-capacitor current to the level needed during the storage state. V_{GM} is defined as the maximum forward bias which can be applied to the gate-to-capacitor or gate-to-bitline junctions before an unacceptably large gate current begins to flow. V_{high} and V_{low} are the logic high and the logic low voltages internal to the memory array. The threshold voltage, V_T , is a fixed gate-to-source value. The logic swing is defined as $V_{high} - V_{low}$.

The cell's write sequences relate the logic swing to the threshold voltage and the maximum gate forward bias. For example, to write a one, the bitline is taken to V_{high} and the transistor is turned on. The transistor must remain on until the capacitor charges up to V_{high} . This requires that the gate (the wordline) be held at $V_{high} + V_T + V_{on}$. The wordline voltage during the write cycle, $V_{WL(on)}$, must not exceed $V_{low} + V_{GM}$ to prevent excessive gate current in other cells connected to the same wordline. Thus:

$$V_{WL(on)} = V_{high} + V_T + V_{on} \leq V_{low} + V_{GM} \quad [4.1]$$

$$V_{high} - V_{low} \leq V_{GM} - V_T - V_{on} \quad [4.2]$$

V_{GM} is established by the access transistor type. Typical values might be 0.6 V, 0.8 V, 1.3 V, and 1.4 V for MESFETs, treated-gate MESFETs, JFETs, and heterojunction JFETs, respectively. V_{on} is selected from the access transistor's I_{DS} -versus- V_{GS} characteristics to satisfy the write cycle time requirements, and must typically be greater than 0.4 V. Equation 4.2 shows that once V_{GM} and V_{on} are established, making the threshold voltage more positive directly reduces the logic swing.

Consideration of the storage state places a minimum value on the logic low voltage. During the storage state, the bitline will be at its precharge level between V_{low} and V_{high} , so the gate-to-source bias will always be more negative than $V_{WL(off)} - V_{low}$. To keep the capacitor-to-bitline leakage at an allowable value, the gate-to-source bias must be more negative than $V_T - V_{off}$, so:

$$V_{WL(off)} - V_{low} \leq V_T - V_{off} \quad [4.3]$$

$$V_{low} \leq V_{WL(off)} + V_{off} - V_T \quad [4.4]$$

The wordline voltage during the off state must be chosen to be the same as the substrate voltage to prevent the large gate-to-substrate punchthrough current described in Section 3.3. Defining the grounded substrate potential as zero,

equation 4.4 becomes:

$$V_{low} \geq V_{off} - V_T \quad [4.5]$$

The value necessary for V_{off} is determined by the subthreshold swing and the acceptable capacitor-to-bitline leakage in the storage state. As described in the previous chapter, the subthreshold swing depends on threshold voltage, gate-to-channel barrier height, channel doping, gate length, and temperature. A conservative estimate of subthreshold ideality factor is 1.4, corresponding to a subthreshold swing of 84 mV/decade at 300 K. Assuming that the leakage current in the storage state must be six orders of magnitude less than at threshold, and that the devices must work at 400 K, a conservative estimate for V_{off} is 0.675 V.

Treating relations 4.1, 4.2, and 4.5 as equalities, V_{low} , V_{high} , and $V_{WL(on)}$ can be plotted as a function of threshold voltage, using V_{on} , V_{off} , and V_{GM} as parameters. Figures 4.1 and 4.2 show these relationships for a MESFET and a JFET. Both plots use $V_{off} = 0.675$ V and $V_{on} = 0.5$ V. For the MESFET, $V_{GM} = 0.6$ V and for the JFET, $V_{GM} = 1.3$ V. In both cases, $V_{WL(off)}$ is required to be zero. For a specified logic swing, there is an upper limit on the threshold voltage for each type of transistor. For $V_{on} = 0.5$ V and a minimum logic swing of 0.5 V, the figures show that the most positive MESFET threshold is -0.4 V, and the most positive JFET threshold is -0.3 V. Similarly, the minimum logic low voltage is 1.075 V for the MESFET and 0.375 for the JFET.

For a specified logic swing, shifting the threshold negatively will allow larger values of V_{low} . This increases capacitor-to-bitline current during reading and writing. However, more negative threshold voltages also require larger values for V_{low} , V_{high} , and $V_{WL(on)}$. $V_{WL(off)}$ must remain at ground, so increasing $V_{WL(on)}$ increases the wordline voltage swing, which will eventually increase the cycle time. The wordline-to-capacitor leakage in the "stored one" state increases with V_{high} . A maximum lower limit on the threshold is reached when the logic high voltage reaches the gate junction breakdown limit. A second-order detrimental effect of more negative threshold voltages is decreased subthreshold slope, as described in the previous chapter.

Summarizing, given a transistor type, signal charge, desired cycle times, and operating temperature range, the values for V_{GM} , V_{on} , and V_{off} are approximately fixed. $V_{WL(off)}$ is required to be zero. The necessary logic swing is fixed by the sense amplifier design, placing an upper limit on the threshold voltage. The exact target for the threshold voltage below this limit is determined by considering the effects of its variation on operating speed and storage time.

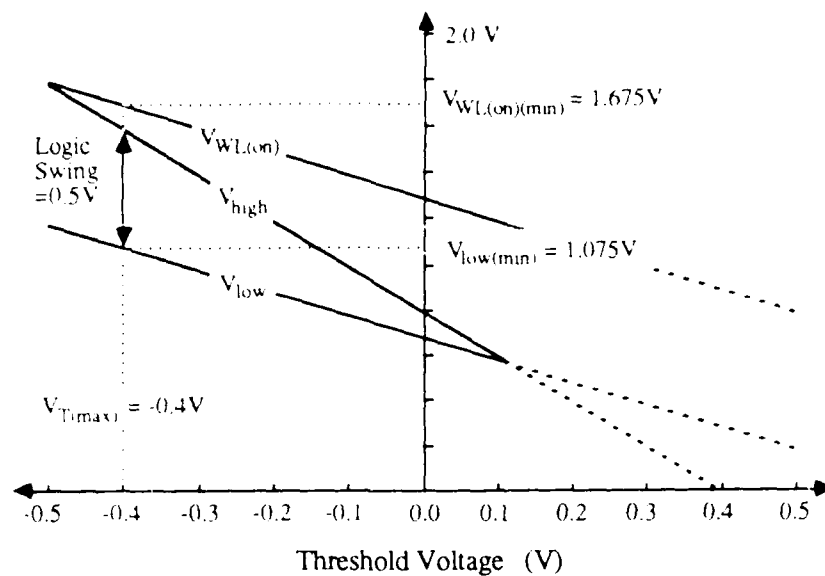


Figure 4.1 Operating voltages versus access-transistor threshold for a MESFET-accessed dynamic memory cell. $V_{GM} = 0.6 \text{ V}$ and $V_{on} = 0.5 \text{ V}$.

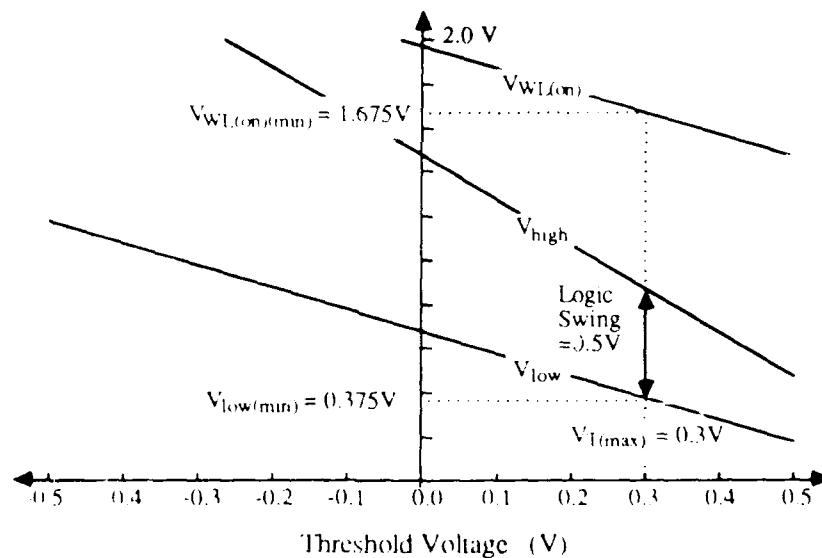


Figure 4.2 Operating voltages versus access-transistor threshold for a JFET-accessed dynamic memory cell. $V_{GM} = 1.3 \text{ V}$ and $V_{on} = 0.5 \text{ V}$.

4.2 Cell Configurations

This section will discuss physical cell configurations for JFET and MESFET direct-access dynamic memory cells. Because the PN-junction capacitor does not require a surface-deposited insulator, there are a variety of conceivable ways to stack capacitors and transistors to reduce the effective cell dimensions. The major disadvantage of such three-dimensional stacking schemes is that parasitic capacitances are greatly increased and speed performance is reduced. As discussed in the first chapter, the targeted applications of the GaAs dynamic memory emphasize ultra-high speed at only moderate density, so three-dimensional configurations which require doped layers directly beneath the access transistor will not be considered here.

The layers underlying the cell will not be included in the diagrams or discussion for simplicity. The short-channel-effect-inhibiting depleted P layer can be added below the channels of the FETs in all of the following designs. AlGaAs layers or low-temperature buffer layers for alpha particle protection and suppression of backgating could be included in the epitaxial designs. It is assumed that all of the cells are constructed on semi-insulating substrates or on doped substrates with a thick (more than 1 μm) undoped epitaxial buffer layer.

Conceptually, the simplest way to connect a field-effect access transistor to a PN-junction capacitor is to allow the channel of the FET to also form the floating N region of the capacitor. An example of such a design using a JFET access transistor is diagramed in Figure 4.3. This design is compact and simple, and it can be constructed without ion implantation using JFET or HJFET access transistors. Such cells have been fabricated to demonstrate the feasibility of FET direct access, but they suffer from a speed limitation which makes them impractical for real applications. When a logic "one" is written to the cell by removing electrons from the floating N region, parasitic transistor action in the capacitor tends to pinch-off the channel nearest to the access transistor, greatly increasing the write time. This parasitic transistor problem points out the need for cell designs in which the transistor and the capacitor can be optimized separately.

One way to separately design the capacitor and channel in epitaxial cells is to form them in different layers and connect them with a metal line. This configuration is shown in Figure 4.4. The cell is somewhat less compact because of the addition of the ohmic contacts and metal line, but the charge storage density of the capacitor is effectively doubled due to the use of both the upper and lower junctions. The capacitance of the lower junction only contributes to the operation of the cell if the lower P⁺ is connected to the upper P⁺ layer. In

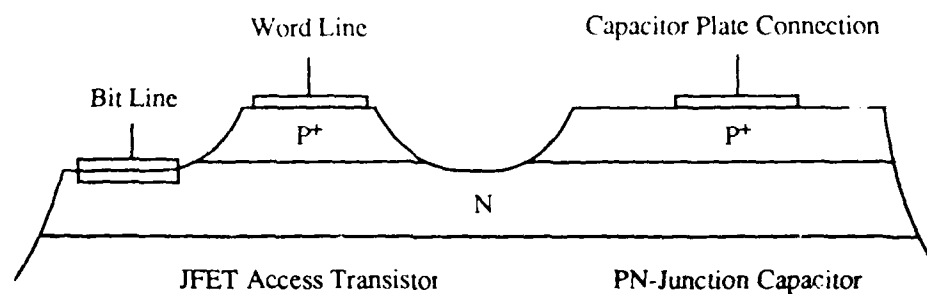


Figure 4.3 A simple epitaxial dynamic memory cell with the transistor channel forming the floating N region of the storage capacitor.

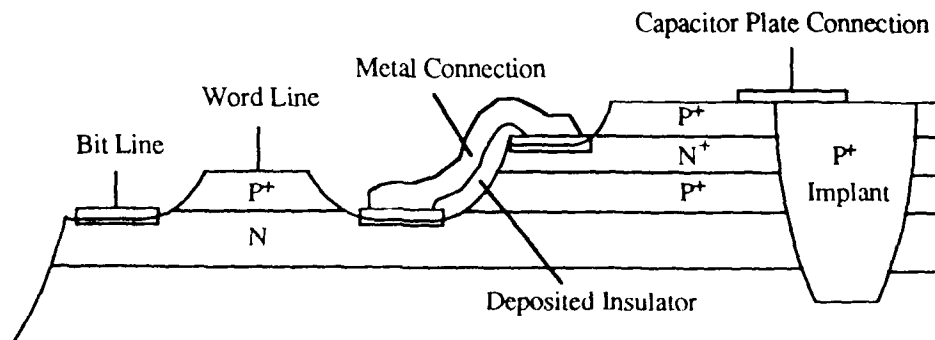


Figure 4.4 An epitaxial cell design with the transistor and capacitor formed from separate layers and connected by a metal line.

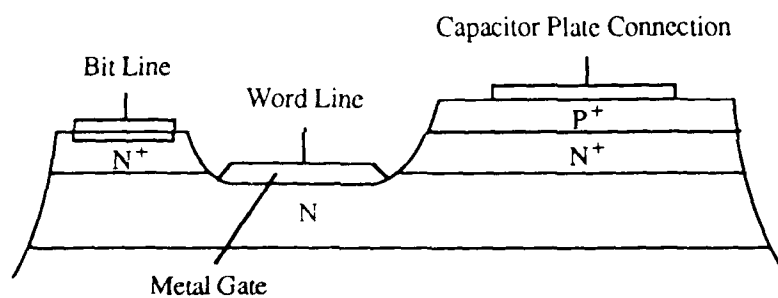
Figure 4.4, both the capacitor isolation and the connection of the upper and lower P^+ plates are accomplished simultaneously by using a deep P^+ implant. The isolation could be done by etching, but then the lower P^+ layer would require an additional ohmic contact.

The addition of the alloyed contacts to the floating N region may increase the number of generation sites and reduce the storage time of the cell. However, test structures with such alloyed contacts have already demonstrated storage times of many seconds at room temperature, as will be reported in the following section.

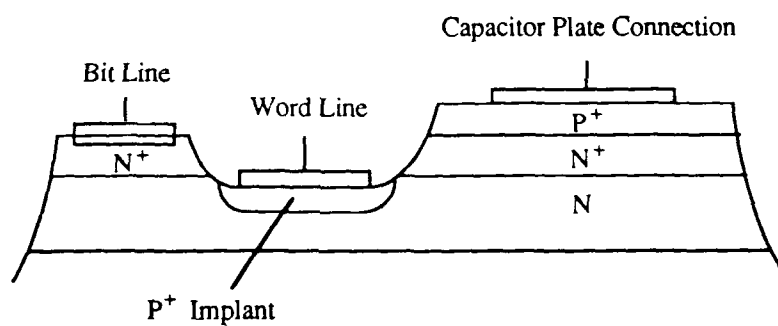
In the cell configuration of Figure 4.4, the transistor and the capacitor are composed of separate layers, so that the transistor choice is unconstrained. The transistor could be an enhancement or depletion JFET, HJFET, MESFET, or even a MODFET. Non-epitaxial-gate FETs, such as MESFETs and implanted JFETs, can also obtain the same independence of threshold voltages in the transistor and capacitor by making use of recessed gates, as is shown in Figure 4.5. In the recessed-gate configurations, the capacitor consists of a single N^+P^+ junction. The N^- layer ensures that the parasitic transistor action which slows devices like Figure 4.3 will not occur. The N^+ layer also allows formation of low resistance bitline contacts. Short-channel epitaxial recess-etched MESFETs of the type shown in Figure 4.5a have demonstrated very high transconductances [56]. Experimental results from this type of device will be presented in the next section.

As shown in Figure 4.5b, it is also possible to ion implant the gate of a recess-etched JFET cell. This cell retains the compactness of the cell shown in 4.3 without the speed limitations. The only penalties incurred compared to the cell of Figure 1.3 are increased generation leakage from the ion-implanted gate and the added complexity of threshold control with both a recess etch and an implanted gate.

All of the cell configurations considered to this point have used epitaxially-grown layers and capacitors with P^+ layers overlying most or all of the floating N regions. It is also possible to design planar, non-epitaxial cells completely produced by ion implantation. Figure 4.6 shows MESFET and JFET accessed versions of all-implanted cells. These cells offer an important advantage beyond planarity and non-epitaxial fabrication. The implanted P layer below the floating N region in the capacitor terminates the field due to the stored charge. The generation volume is restricted, so that the defects and generation sites in the undoped material below the capacitor cannot limit the storage time. Perhaps more importantly, the collection volume for ionized carriers created

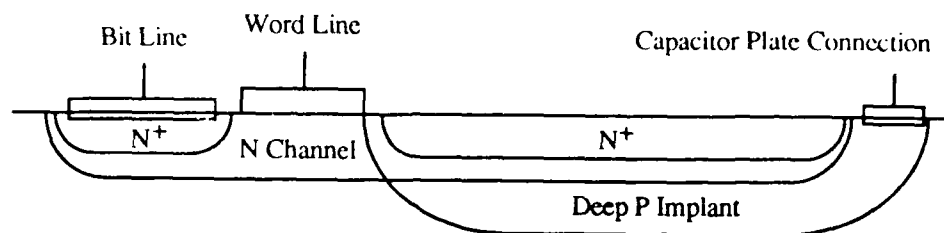


a) MESFET version

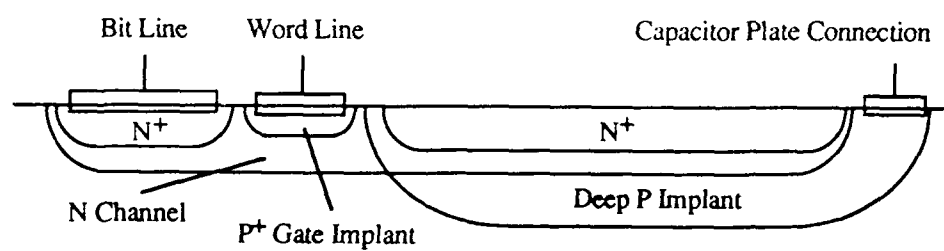


b) Ion-implanted JFET version

Figure 4.5 Dynamic memory cells using recess-etched gates.



a) MESFET version



b) JFET version

Figure 4.6 Planar ion-implanted dynamic memory cell configurations.

during an alpha particle strike is greatly reduced, resulting in improved immunity from soft errors.

Additionally, the charge storage density in the all-implanted cells could be increased by the addition of a deposited insulator and grounded metal or polysilicon plate above the floating N layer, as shown in Figure 4.7. The MIS capacitor is not required to exhibit accumulation/depletion/inversion behavior, because the N^+ region still provides majority carriers to form the signal charge. The only requirements on the MIS capacitor are that it does not introduce large leakage currents and that its fabrication sequence is compatible with GaAs processing. Such an MIS capacitor has already been demonstrated using Si_3N_4 as an insulator for use in capacitor-FET logic for GaAs static memories [57]. The cell configuration in Figure 4.6 is identical in form and function to the grounded-plate Hi-C Si dynamic memory cell design [58], with the exception of the MESFET or JFET access transistor replacing the Si MOSFET.

All of the cell configurations shown in Figures 4.4 through 4.7 have features which could recommend them for use in certain applications. The metal-connected JFET and HJFET cells offer epitaxial transistors and high charge capacity per unit area. The recess-etched structures offer a very compact cell design with high-speed transistors. The ion-implanted structures offer planar design, independence from epitaxy, and improved alpha-particle immunity. The major individual components of all these designs have already been demonstrated, as shown in Chapter 2 and in the following section.

4.3 Experimental Demonstration of Complete Dynamic Memory Cells

The first demonstration of a complete GaAs dynamic memory cell with a directly-connected access transistor used the configuration shown in Figure 4.3. The channel was 2500 Å thick with a doping of $1.2 \times 10^{17} \text{ cm}^{-3}$, resulting in a threshold voltage near -1 V. The 1.5 μm thick epitaxial layer under the channel was doped to approximately 10^{16} cm^{-3} P type, on a P^+ substrate. Ohmic bitline contacts were made with alloyed AuGe. The wordline and capacitor plate contacts were electron-beam evaporated Ti/Au.

The storage time of the cell was tested by applying a bias pulse directly to the capacitor plate to dump electrons from the N region. The capacitance-recovery transient was observed with the bitline and wordline floating, and with the bitline grounded and the transistor turned on and off. The capacitance-recovery transients for these three conditions are shown in Figure 4.8. When the bitline and wordline are both floating, a storage time of about 5 seconds is

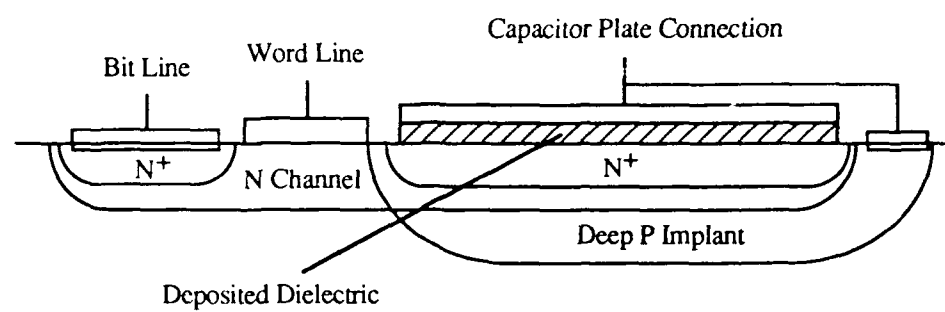


Figure 4.7 Ion-implanted cell with MIS capacitor for increased charge-storage density.

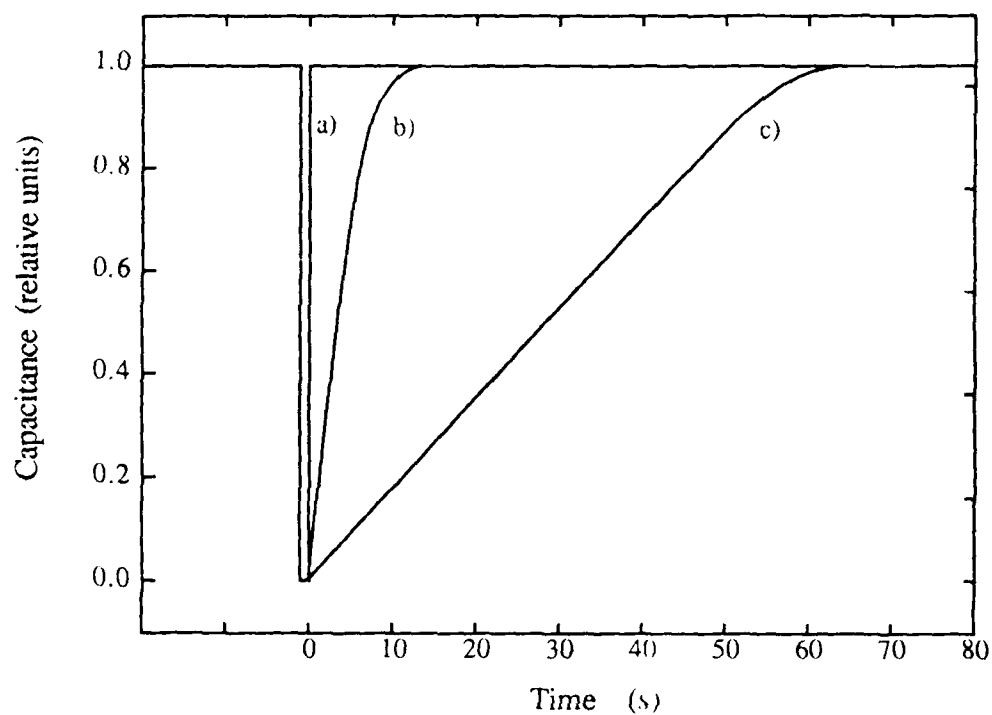


Figure 4.8 Capacitance-recovery transients for a JFET-isolated cell with a doped buffer layer. A one-second dump pulse is applied to the capacitor just prior to $t = 0$.

- a) bitline grounded, transistor on
- b) bitline and wordline floating
- c) bitline grounded, transistor off

observed at room temperature. With the bitline grounded and the transistor turned on, the storage well is immediately refilled with electrons, and the recovery transient is abrupt. With the bitline grounded and the transistor turned off, the storage time increases to about 40 seconds. The increase in storage time from the case with the bitline and wordline floating to the case with the bitline grounded and the transistor turned off indicates that leakage from generation around the alloyed bitline contact is larger than the subthreshold transistor leakage. The linearity of the recovery transient when the transistor is turned off suggests that the storage time is being limited by transistor leakages, because the transistor leakage currents are less sensitive to the potential of the floating N region.

It is not practical in large dynamic memory arrays to have an undepleted doped layer lying under the access transistor and bitline contact because of the parasitic capacitances introduced. The simple JFET-isolated configuration of Figure 4.3 was also fabricated from a MBE-grown film with dimensions and dopings as shown in Figure 4.9 [59]. The thick 10^{16} cm^{-3} doped buffer layer used in the first set of devices was replaced by a completely-depleted 500 Å P layer doped at 10^{17} cm^{-3} and a 1 μm undoped buffer. 5000 Å of a P^+ material was grown below the intrinsic buffer to screen any generation centers at the substrate interface. The channel doping was increased to $4 \times 10^{17} \text{ cm}^{-3}$, and its thickness was reduced to 1300 Å, maintaining a threshold voltage near -1 V. The intrinsic buffer layer results in a large generation volume, and the storage time of ungated test capacitors was only a few seconds at room temperature. The implications of the reduced storage times and possible remedies will be discussed in the next chapter.

A surface view of the cell layout is shown in Figure 4.10. Two separate capacitor plates are included to allow surface-to-surface capacitance measurement, because the intrinsic buffer layer makes the surface-to-substrate capacitance very small. The ring gate is intended to separate the storage region from the isolation-mesa edge, and to allow the gate to remain entirely on the mesa. The additional alloyed contact to the N region allows direct measurement of the transistor characteristics. The area of the storage capacitors is $6 \times 10^4 \mu\text{m}^2$. The inside perimeter of the gate is 1120 μm .

Figures 4.11 and 4.12 show digitized copies of experimental data illustrating read and write sequences for the cell. In Figure 4.11, information is written to the cell through the access transistor while the charge state of the cell is monitored through its capacitance. A high or low potential on the bitline is gated to the cell during the narrow gate pulses which briefly turn the transistor on. If the bitline is at the higher potential when the transistor is turned on,

Gate and Capacitor Plate	$N_A = 10^{19} \text{ cm}^{-3}$	500 Å
Channel and Floating N Region	$N_D = 4 \times 10^{17} \text{ cm}^{-3}$	1300 Å
Depleted P Layer	$N_A = 10^{17} \text{ cm}^{-3}$	500 Å
Undoped Buffer Layer		1.0 μm
P ⁺ Layer		5000 Å
P ⁺ Substrate		

Figure 4.9 Dopings and dimensions MBE film for intrinsic-buffer JFET-accessed dynamic memory cells.

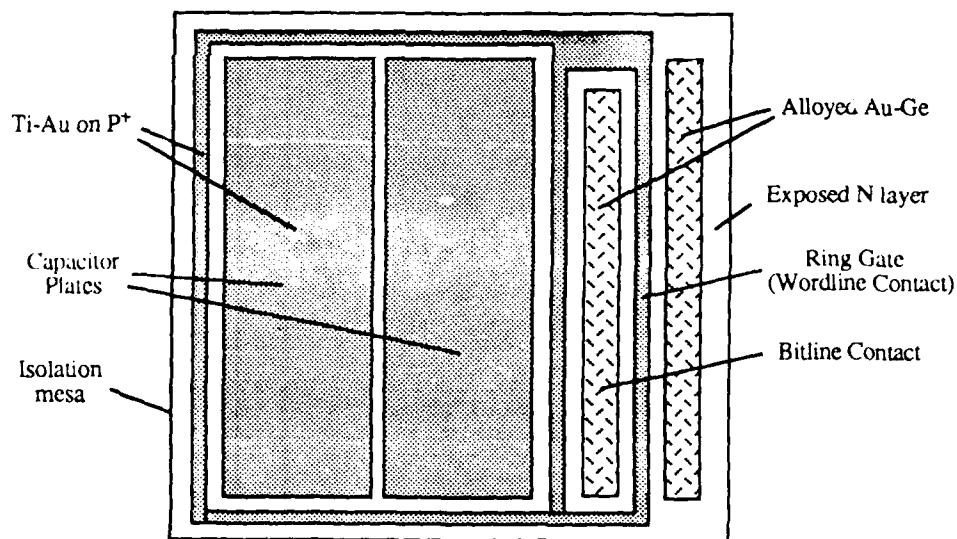


Figure 4.10 Layout of JFET-accessed ring-gate test device.

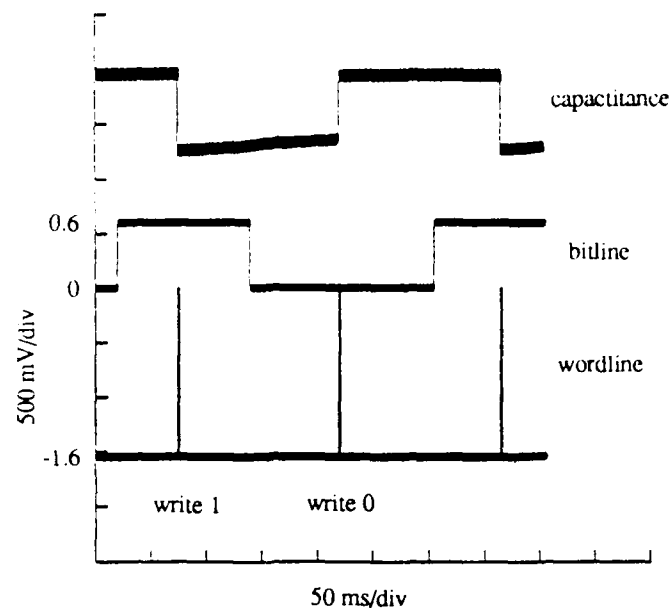


Figure 4.11 Demonstration of writing the JFET-accessed cells through the transistor.

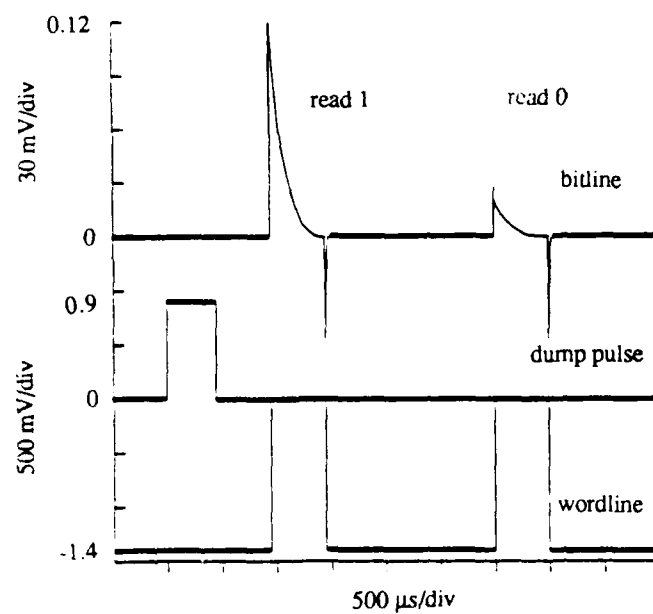


Figure 4.12 Demonstration of reading the JFET-accessed cells through the transistor.

electrons are extracted from the well, and the capacitance is reduced. When the gate bias is returned to -1.6 V, the transistor is shut off, and the capacitor retains a positive charge. The small upward transient during the "store 1" state is due to leakages. When the bitline is returned to ground and the transistor is turned on, the capacitor is immediately restored to its equilibrium state.

Figure 4.12 demonstrates reading the cell through the access transistor. The dump pulse is applied directly to the capacitor plate contacts, and the upper trace is the output of a low-capacitance active probe monitoring the potential of the bitline contact. When the transistor is turned on after electrons have been removed from the floating N region, the bitline jumps to a positive potential as electrons are pulled from bitline to capacitor. When the transistor is turned on with no prior dump pulse, the bitline jumps to a much lower potential as carriers redistribute under the gate depletion region. The narrow negative spikes when the gate turns off are also due to gate depletion region redistribution. The decay of the bitline potential while the transistor is on is caused by charge leaking through the finite input impedance of the active probe. The sense amplifier must distinguish between the initial bitline potentials corresponding to a stored 1 and a stored 0.

Figure 4.13 shows the capacitance-recovery time constant of a JFET-accessed cell versus temperature. The measurement was made by gating a 0.5 V bitline bias to the cell and then returning the gate to -1.7 V and observing the capacitance-recovery transient, with the capacitor plates and the substrate held at ground. The storage time decreases exponentially, reaching 100 ms at 70° C. The activation energy is 0.63 eV.

The write-and-store sequence illustrated in Figure 4.11 uses both positive and negative power supplies. The wordline is held at a large negative bias with respect to the grounded substrate during storage. This means that during the storage state, the large gate-to-substrate punchthrough current documented in Section 3.3 will be present. To avoid the punchthrough current and to require only a positive power supply, the logic low voltage must be shifted positively from zero, as described in Section 4.1. Figure 4.14 illustrates the write-and-storage sequence for the same JFET-accessed device using realistic operating voltages. The bitline trace shows that the logic low value has been shifted to $+1.5$ V, and the logic high has become $+2.2$ V. The wordline is held at ground during the storage state and goes to $+2.3$ V to write information to the cell. The capacitor plate contact is held at the logic-low voltage of 1.5 V. To hold both capacitor plates at a DC bias, the measurement set-up requires the capacitance to be monitored from surface-to-substrate, rather than from surface-to-surface. This results in a very small signal which is easily affected by

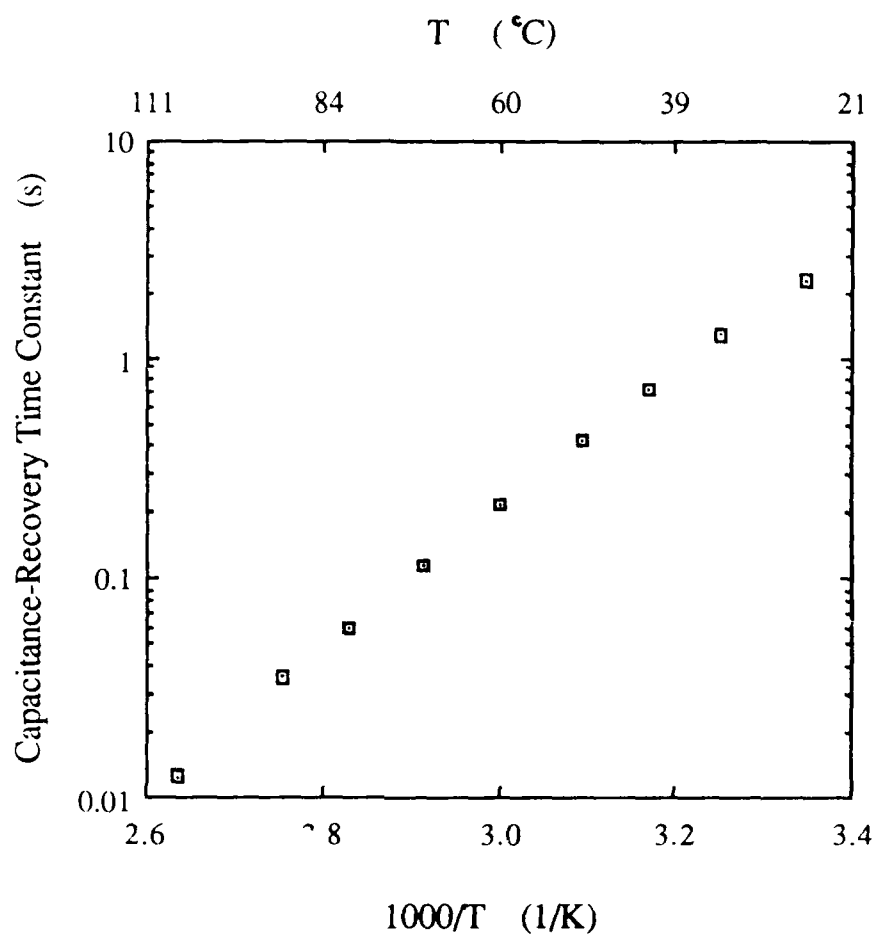


Figure 4.13 Capacitance-recovery time constant versus temperature for a ring-gate JFET-accessed memory cell.

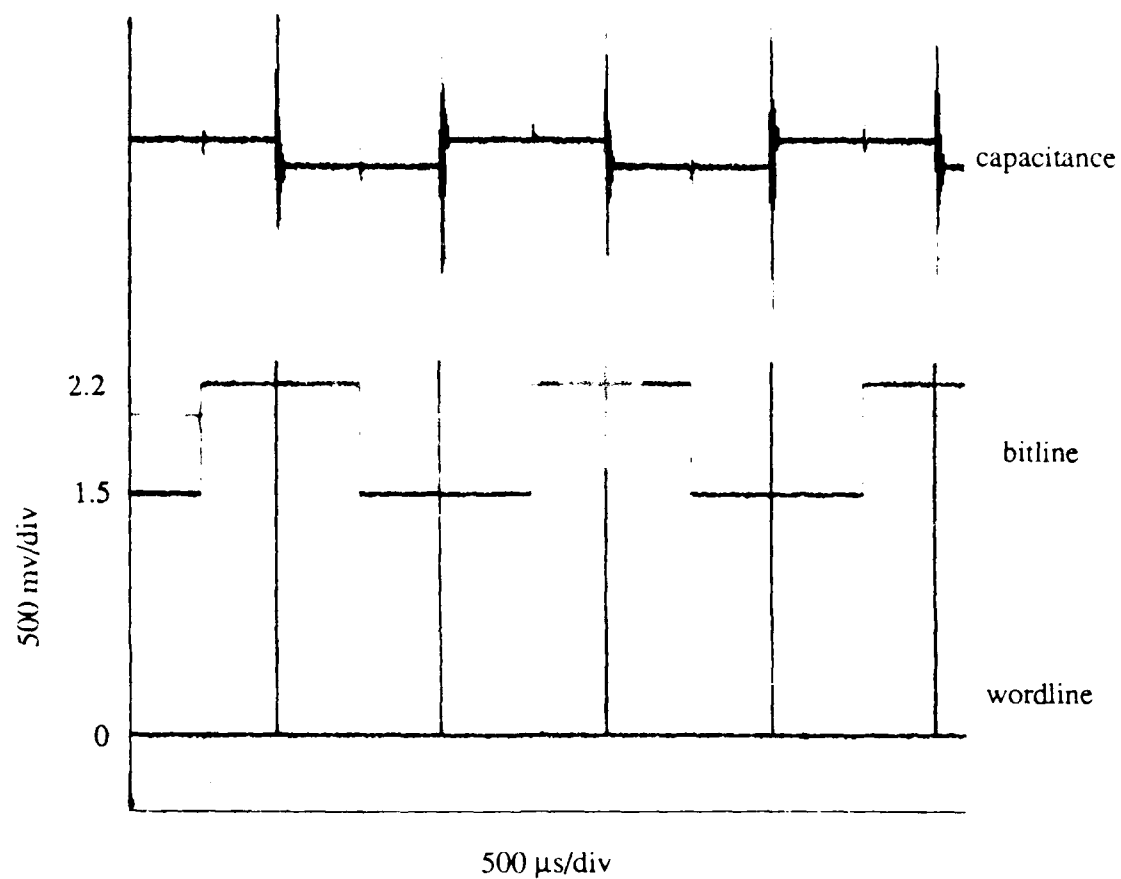


Figure 4.14 Write-and-storage sequence for a JFET-accessed cell using positively-shifted logic levels.

ringing and capacitively-coupled noise, but the actual stored charge is not reduced compared to the conditions of Figure 4.11.

As mentioned in the previous section, the simple JFET-accessed configuration of Figure 4.3 is not a practical cell for a high-speed dynamic memory because parasitic transistor action in the capacitor greatly increases the time needed to pull electrons from the storage region. For example, when the narrow gate pulses in Figure 4.11 or 4.14 are reduced to less than about 10 μ s, the response of the capacitor starts to decrease. SPICE simulations confirm that this is the result of the capacitor's N region pinching off near the access transistor, creating a voltage drop which impedes the removal of the remaining electrons [60].

Recessed-gate MESFET accessed cells of the type shown in Figure 4.5a have also been demonstrated. For the fabrication of the MESFET devices, the JFET film design shown in Figure 4.9 was modified by the insertion of a 1000 Å N⁺ layer and a 300 Å intrinsic buffer layer between the channel and the P⁺ layer on top. The channel layer thickness was also increased to 1500 Å, but the final channel thickness is determined by the depth of the recess etch. The gates consist of electron-beam evaporated Au-Ti-Au. Gold was chosen as the metal to contact the GaAs because it has demonstrated increased Schottky-barrier heights using the (NH₄)₂S treatment. The complete cells which have been fabricated to date, and which are reported here, use untreated gates. The gold gates will allow direct comparison with future devices employing the (NH₄)₂S treatment.

Figure 4.15 shows a write-and-storage sequence for a MESFET-accessed device at room temperature. The cell has a ring gate layout similar to Figure 4.10. The write-and-storage sequence of Figure 4.15 uses positively-shifted logic voltages so that the wordline is at ground during storage, eliminating gate-to-substrate punchthrough current. The N⁺ layer in the recess-etched MESFET structure reduces the wordline contact resistance and eliminates the parasitic transistor action which slowed the "write 1" operation in the JFET-accessed cells. The width of the wordline pulses used to write information to the cell could be reduced to less than 100 ns, which was the limit of the measurement set-up, without loss of functionality.

The longest storage time measured for these untreated MESFET-accessed devices is only about 80 ms at room temperature. However, the activation energy of the storage-time-limiting gate leakage is very low, typically less than 0.5 eV. The minimum drain current achieved by a MESFET access transistor increases by only an order of magnitude from room temperature to 100 °C. If

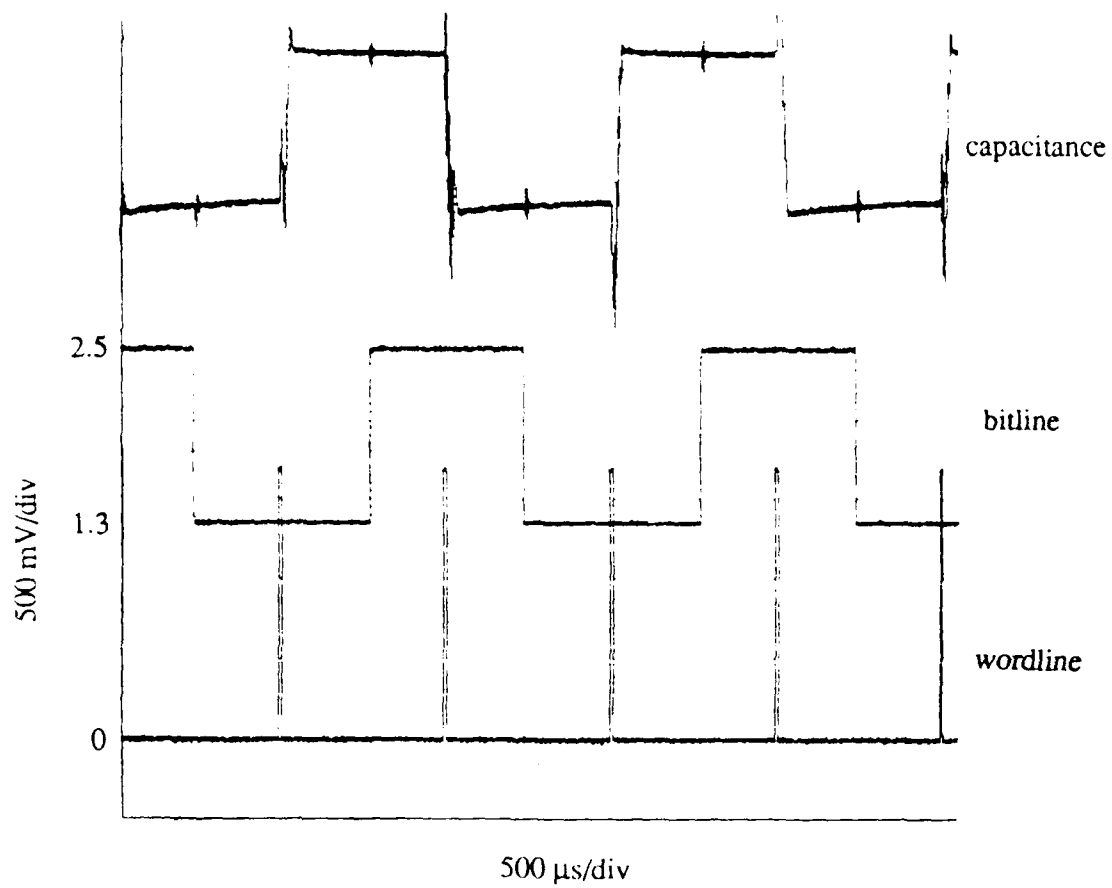


Figure 4.15 Write-and-storage sequence for a MESFET-accessed cell using positively-shifted logic levels.

the tenfold decrease in the current observed with $(\text{NH}_4)_2\text{S}$ -treated gold Schottky barriers can be achieved in the access transistor gates, MESFET-accessed devices should be capable of operating well above room temperature.

CHAPTER 5

RECOMMENDATIONS FOR FUTURE RESEARCH

5.0 Introduction

Several important existence proofs of key features necessary for a useful GaAs dynamic memory technology have been reported in the previous chapters. Individual etch-isolated storage capacitors have been demonstrated which simultaneously provide high charge-storage density and sufficient storage times to allow operation at high temperatures ($>125^{\circ}\text{C}$) even with minimum-sized devices ($<4\times4\mu\text{m}^2$). An ion-implanted capacitor has been demonstrated with a storage time of greater than 100 ms up to 70°C . All-epitaxial JFET-accessed complete memory cells have also shown 100 ms storage times up to 70°C . Untreated recessed-gate MESFET cells have been demonstrated at room temperature with very short read and write time requirements. The remaining challenge is to put the individual pieces together to produce complete cells that are at once fast, compact, and capable of working at high temperatures.

5.1 Improving the Experimental Cells

The two complete cell designs which have presently been demonstrated suffer from two individual problems and one common problem. The demonstrated all-epitaxial JFET-accessed cells are not realistically usable because of the long write times caused by parasitic transistor action in the capacitor. The experimental MESFET-accessed devices have insufficient storage time for operation above room temperature due to large gate leakage over the small gate-to-channel barrier. The common problem of both cells is the short storage times caused by the large generation volume in the undoped buffer regions below the storage capacitors.

The parasitic transistor action in the JFET cell can be eliminated by switching from the present all-epitaxial configuration (Figure 4.3) to a combination of epitaxy and ion implantation as in the recess-gate cell of Figure 4.5b. The leakage from the implanted gate should be low enough to permit

excellent storage times if the results reported in reference 47 can be duplicated. The large gate leakage of the MESFET-accessed cell should be reduced by the ammonium-sulfide Schottky barrier treatment of reference 19.

The precise reason for the reduced storage times in the capacitors of the complete memory cells is not presently understood. The very small thickness of the upper P^+ layer could be somehow reducing the storage times, but this doesn't appear to be the case, because the thickness was increased from 500Å in the JFET cells to 700Å in the MESFET cells with no corresponding performance improvement. Theoretically, the storage time should scale with the doping in lightly-doped devices where field-enhanced generation is not significant. Comparison with results from Section 2.2.5 shows that to explain the observed results, the background doping in the unintentionally-doped buffer layer would have to be lower than $2 \times 10^{14} \text{ cm}^{-3}$, which is a reasonable number for high-quality MBE-grown GaAs. However, a confusing experimental detail about the capacitors with undoped buffer layers is the relatively low activation energy, 0.63 eV, measured for the JFET-accessed cells as shown in Figure 4.13. If the limiting generation is coming from the lightly-doped lower junction, perimeter generation should dominate and the activation energy should be 0.79 eV.

To build useful versions of the recessed-gate cells shown in Figure 4.5, the short storage time of the capacitors on undoped buffer layers must be improved or at least understood. Three experiments which might provide useful information are:

- 1) Using a mask set with capacitors all of large area but with widely varying perimeters, measure storage time performance versus perimeter-to-area ratio. This should determine if the devices are being limited by perimeter generation or by generation at defect sites in the bulk. The large area of the test capacitors must be maintained because of the low surface-to-substrate capacitance of the devices.
- 2) Check the effects of a thin upper P^+ layer by thinning existing $P^+N^+P^+$ or $P^+iN^+iP^+$ films, or by growing a new film identical to the JFET film of Figure 4.9 but with the undoped buffer removed.
- 3) Attempt to reduce the surface generation in the undoped layer by using an AlGaAs layer under the channel. The present design uses a shallow mesa etch to isolate the capacitors which results in a large surface-generation area as shown in Figure 5.1. An AlGaAs layer inserted under the channel could reduce the surface generation and improve the storage time, as was observed for the buried-well devices of Figure 2.15. An AlGaAs layer could also act as an etch stop to minimize the isolation-mesa step size, as diagramed in Figure 5.2, and an AlGaAs layer under

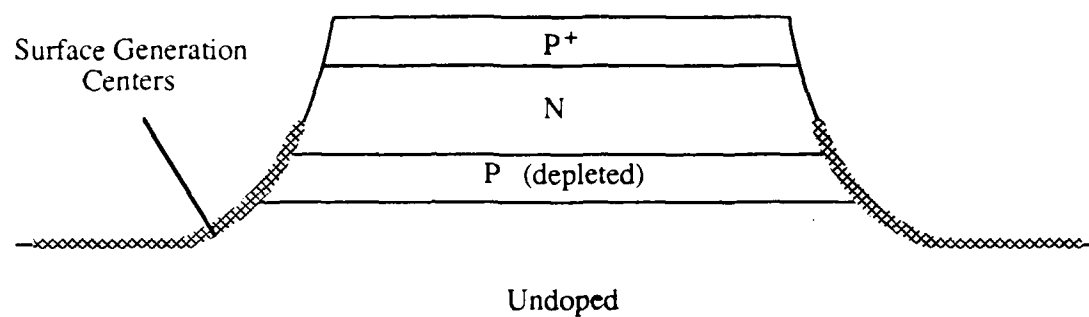


Figure 5.1 Dynamic memory cell capacitor on an undoped buffer layer showing surface-generation sites.

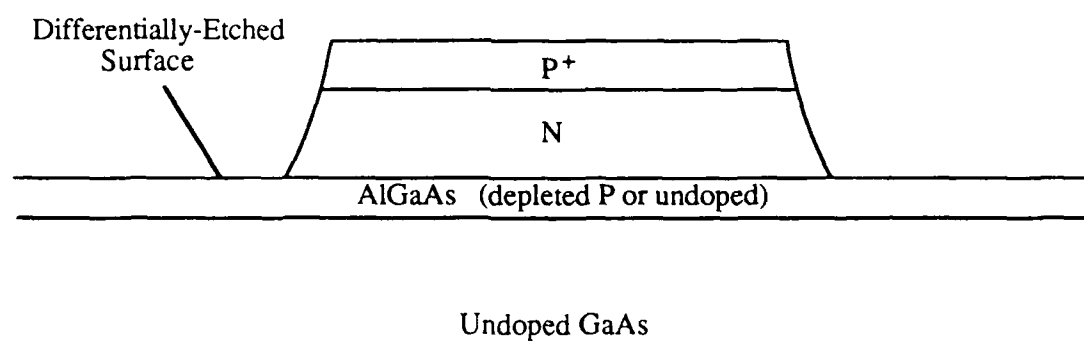


Figure 5.2 Dynamic memory cell capacitor with an $AlGaAs$ layer as an etch stop and surface-passivation layer.

the channel has been shown to be beneficial in suppressing short channel effects in FETs [54].

Even if the generation rate can be reduced so that scaling of the capacitors to small dimensions is possible, the problem of high alpha-particle collection efficiency will remain. This problem will have to be combatted by maintaining a robust signal charge and reducing radiation sources to a minimum.

5.2 Continued Ion-Implantation Research

The ion-implanted cell configurations of Figures 4.6 and 4.7 are attractive for several reasons. They have a cost advantage because they require no epitaxy. They are planar for improved manufacturability. Perhaps most importantly, the grounded P region under the N region in the capacitor results in a very small generation volume compared to the cells described in the last section. This means that alpha-particle immunity should be greatly improved, and that only those defects within the implanted region can contribute to generation.

The ion-implanted capacitor results reported so far have been encouraging, but much work remains to be done. The relative importance of the perimeter and bulk generation components in ion-implanted capacitors must be established to predict scaling effects on storage time. Continued efforts are needed to understand the effects of implanting and annealing conditions and procedures on storage-time performance. Specifically, experimental correlations of activation percentages in the P and N layers with storage time are needed to assess how much more performance improvement can be expected.

The Mg and Si implantation experiments described in Section 2.2.7.2 used a P^+ layer implanted over an N layer, which is the opposite order required for the cells shown in Figures 4.6 and 4.7. The effects of reversing the order of the implants needs to be experimentally examined. Complete ion-implanted cells need to be designed, fabricated, and tested.

If the leakage rate in the ion-implanted capacitors cannot be reduced sufficiently to allow satisfactory scaling, then a hybrid structure using epitaxy and implantation might be beneficial. Such a device using an implanted FET with an epitaxial capacitor is shown in Figure 5.3. The area of the implanted transistor contact can be minimized while the epitaxial capacitor provides a large charge storage capacity. The storage time might be improved by as much as an order of magnitude over a design in which the entire capacitor is formed by implantation.

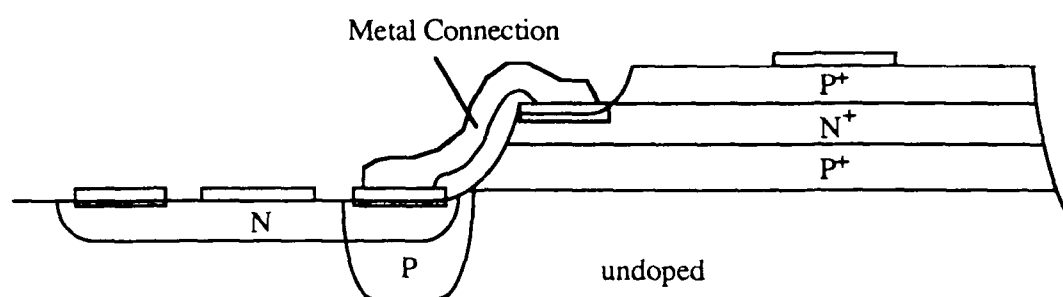


Figure 5.3 A hybrid cell using an implanted transistor and an epitaxial capacitor.

5.4 Experimental Investigation of Transistor Scaling Effects

The perimeter or area dependence of the gate leakage current in both the MESFETs and the JFETs needs to be experimentally established. If the gate leakage scales with the area of the gate, then improved performance can be obtained by reducing the gate length until short channel effects make bitline-to-capacitor leakage the limiting factor. The capacitor-area-to-gate-area ratio in the experimental ring-gate MESFET devices is about 3.75:1. In a cell with a $10 \times 10 \mu\text{m}^2$ capacitor and a $10 \times 1 \mu\text{m}^2$ gate, the ratio would be 10:1, so the present experimental device is a worst-case test of gate leakage if the leakage scales with area. However, the ratio of capacitor area to the inner perimeter of the gate in the experimental devices is about $50:1 \mu\text{m}$. In the $10 \times 10 \mu\text{m}^2$ cell this ratio would be only 10:1, so if the gate leakage scales with the gate perimeter, the performance will decrease with decreasing dimensions. The performance of cells with different capacitor-area-to-gate-area and capacitor-area-to-gate-perimeter ratios needs to be experimentally established.

5.5 Circuit Development

Section 4.1 began a discussion of circuit requirements for FET-accessed dynamic memory cells in GaAs by considering operating voltages internal to the memory array. The design of the peripheral circuits is critical to the performance of the memory. Circuits must be designed to provide the internal supply voltage to hold the capacitor plate connections at the shifted logic-low value, and to drive the wordline voltage from ground to the relatively high value needed for cell access. To maintain very low power consumption it is important that the wordline and substrate potentials be as close as possible during the storage state. Sense amplifiers must be designed to precharge the bitlines to the proper value between the shifted logic-high and logic-low voltages. The sense amplifiers must also transform the shifted internal logic voltages back to the standard DCFL logic levels used by the peripheral circuitry. The speed and sensitivity of the sense amplifiers will largely determine how fast the memory will operate and how small the capacitors can become.

In order to design the peripheral circuitry, computer models which accurately describe the experimental behavior of the test devices must be developed and maintained. Computer simulations are needed to optimize the selection of operating voltages as outlined in Section 4.1. The relationships between the physical design parameters and the performance of individual memory cells are becoming increasingly well established. The optimal selection

of many of the most critical of those parameters, such as channel depth (threshold voltage), gate length and width, and capacitor size now depend on accurately assessing the requirements of the complete dynamic memory circuit.

LIST OF REFERENCES

LIST OF REFERENCES

- [1] J. Miyamoto, S. Saito, H. Mosose, H. Shibata, K. Kanzaki, and T. Iizuka, "A High-Speed 64K CMOS RAM with Bipolar Sense Amplifiers," IEEE Journal of Solid-State Circuits, vol. 19, p. 557, 1984
- [2] A. Cho, "Recent Advances in GaAs on Si," Proceedings of IEDM, p. 901, 1987
- [3] S. Miyazawa and Y. Ishii, "Dislocations as the Origin of Threshold Voltage Scatterings for GaAs MESFETs on LEC-Grown Semi-Insulating GaAs Substrates," IEEE Transactions on Electron Devices, vol. 31, no. 8, p. 1057, 1984
- [4] M. Simmons, "Radiation Effects in GaAs Integrated Circuits: A Comparison with Si," Proceedings of IEEE GaAs IC Symposium, Oct. 1983
- [5] L. Eastman, "Semi-Insulating GaAs Substrates for Integrated Circuit Devices: Promises and Problems," Journal of Vacuum Science and Technology, vol. 16, no. 6, p. 2050, 1979
- [6] W. Helbig and V. Multinovic, "A DCFL E/D MESFET GaAs Experimental RISC Machine," IEEE Tutorial on GaAs Computer Design, p. 339, 1988
- [7] P. Solomon and H. Morkoc, "Modulation-Doped GaAs/AlGaAs Heterojunction Field-Effect Transistors (MODFETs), Ultrahigh-Speed Device for Supercomputers," Transactions on Electron Devices, vol. 31, no. 8, p. 1015, 1984
- [8] H. Kroemer, "Heterostructure Bipolar Transistors and Integrated Circuits," IEEE Proceedings, vol. 70, no. 1, p. 13, 1982
- [9] M. Wilson, Y. Shen, and B. Welch, J. Lee, R. McCullough, J. Salerno, and J. Fan, "GaAs-on-Si: A Manufacturer's Perspective," GaAs IC Symposium Technical Digest, p. 243, 1988

- [10] D. Hodges and H. Jackson, *Analysis and Design of Digital Integrated Circuits*, p. 378, McGraw-Hill, 1983
- [11] H. Makino, S. Matsue, M. Noda, N. Tanino, S. Takano, K. Nisitani, and S. Kayano, "A 7ns/850mW GaAs 4Kb SRAM Fully Operative at 75 ° C," GaAs IC Symposium Technical Digest, p. 71, 1988
- [12] W. Terrell, C. Ho, and R. Hinds, "Direct Replacement of Silicon ECL and TTL SRAMs with High-Performance GaAs Devices," GaAs IC Symposium Technical Digest, p. 79, 1988
- [13] T. Mimura, N. Yokoyama, and M. Fukuta, "Electrical Characteristics of the Plasma-Grown Native-Oxide/GaAs Interface," *Applied Physics Letters*, vol. 34, no. 10, p. 642, 1979
- [14] T. Mimura, and M. Fukuta, "Status of the GaAs Metal-Oxide-Semiconductor Technology," *IEEE Transactions on Electron Devices*, vol. 27, no. 6, p. 1147, 1980
- [15] W. Spicer, I. Lindau, p. Skeath, C. Su, and P. Chye, "Unified Mechanisms for Schottky Barrier Formation and III-V Oxide Interface States," *Physical Review Letters*, vol. 44, p. 420, 1980
- [16] S. Offsey, J. Woodall, A. Warren, P. Kirchner, T. Chappel, and G. Pettit, "Unpinned (100) GaAs Surfaces in Air Using Photochemistry," *Applied Physics Letters*, vol. 48, p. 475, 1986
- [17] C. Sandroff, R. Nottenburg, J. Bischoff, and R. Bhat, "Dramatic Enhancement in the Gain of a GaAs/AlGaAs Heterostructure Bipolar Transistor by Surface Chemical Passivation," *Applied Physics Letters*, vol. 51, p. 33, 1987
- [18] E. Yablonovitch, C. Sandroff, R. Bhat, and T. Gmitter, "Nearly Ideal Electronic Properties of Sulfide-Coated GaAs Surfaces," *Applied Physics Letters*, vol. 51, p. 439, 1987
- [19] M. Carpenter, M. Melloch, and T. Dungan, "Schottky Barrier Formation on (NH₄)₂S Treated N and P Type (100) GaAs," *Applied Physics Letters*, vol. 53, no. 1, p. 66, 1988
- [20] Q-D. Qian, M. Melloch, and J. Cooper, Jr., "Photosensitive Capacitance-Voltage Characteristics of Molecular Beam Epitaxially Grown GaAs/AlGaAs/GaAs Heterostructures," *Applied Physics Letters*, vol. 48, no. 10, p. 638, 1986

- [21] J. Cooper, Jr., Q-D. Qian, and M. Melloch, "Evidence of Long-Term Storage of Minority Carriers in N^+ -GaAs/AlGaAs/ P^- GaAs MIS Capacitors," *IEEE Electron Device Letters*, vol. 7, no. 6, p. 374, 1986
- [22] M. Melloch, J. Cooper, Jr., and Q-D. Qian, "Investigation of Minority Carrier Retention Behind AlAs Barriers," *Journal of Vacuum Science and Technology*, vol. B5, no. 3, p. 766, 1987
- [23] T. Lee, W. Holder, and A. Cho, "AlGaAs-GaAs Double-Heterostructure Small-Area Light Emitting Diodes by MBE," *Applied Physics Letters*, vol. 32, p. 415, 1978
- [24] F. Capasso, F. Beltram, R. Malik, and J.F. Walker, "New Floating-Gate AlGaAs/GaAs Memory Devices with Graded-Gap Electron Injector and Long Retention Times," *Electron Device Letters*, vol. 9, no. 8, p. 377, 1988
- [25] J. Kleine, M. Melloch, and J. Cooper, Jr., "The Effect of Al Mole Fraction on Electron Emission at the $Al_xGa_{1-x}As/GaAs$ Interface," submitted for publication, 1989
- [26] Q. Qian, J. Qui, M. Melloch, J. Cooper, Jr., L. Kolodziejski, M. Kobayashi, and R. Gunshor, "Low Interface State Density at an Epitaxial ZnSe/Epitaxial GaAs Interface," *Applied Physics Letters*, vol. 54, no. 14, 1989
- [27] G. Fountain, S. Hattangady, D. Vitkavage, R. Rudder, and R. Markunas, "GaAs MIS Structures with SiO_2 Using a Thin Si Interlayer," *Electronics Letters*, vol. 24, no. 18, p. 1134, 1988
- [28] R.F. Pierret, *Advanced Semiconductor Fundamentals*, Modular Series on Solid State Devices, vol. 4, p. 117, Addison-Wesley Publishing Company, 1987
- [29] L. Lewyn and J. Meindl, "Physical Limits of VLSI dRAMs," *IEEE Transactions on Electron Devices*, vol. 32, no. 2, p. 311, 1985
- [30] P. Chatterjee, G. Taylor, R. Easley, H-S Fu, and A. Tasch, Jr., "A Survey of High-Density Dynamic RAM Cell Concepts," *IEEE Transactions on Electron Devices*, vol. 26, no. 6, p. 827, 1979
- [31] R.F. Pierret, *Advanced Semiconductor Fundamentals*, Modular Series on Solid State Devices, vol. 4, p. 139, Addison-Wesley Publishing Company, 1987

- [32] W. Shockley, "The Theory of p-n Junctions in Semiconductors and p-n Junction Transistors," Bell System Technical Journal, vol. 28, p. 435, 1949
- [33] C. Henry, R. Logan, and F. Merritt, "The Effect of Surface Recombination Current in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ Heterojunctions," Journal of Applied Physics, vol. 49, no. 6, p. 3530, 1978
- [34] T. Dungan, J. Cooper, Jr., and M. Melloch, "A Thermal-Generation-Limited Buried-Well Structure for Room-Temperature GaAs Dynamic RAMs," IEEE Electron Device Letters, vol. 8, no. 5, p. 243, 1987
- [35] T. Dungan, J. Cooper, Jr., and M. Melloch, "Room-Temperature Dynamic Memories for GaAs Integrated Circuits," Proceedings of IEDM, p. 348, 1987
- [36] H. Poole, "On the Electrical Conductivity of Some Dielectrics," Philosophical Magazine, vol. 42, p. 488, 1921
- [37] J. Frenkel, "On Pre-Breakdown Phenomena in Insulators and Electronic Semi-Conductors," Physical Review, vol. 54, p. 647, 1938
- [38] A. Tsch, Jr. and C. Sah, "Recombination-Generation and Optical Properties of Gold Acceptor in Silicon," Physical Review B, vol. 1, no. 2, p. 800, 1970
- [39] L. Rosier and C. Sah, "Thermal Emission and Capture of Electrons at Sulphur Center in Silicon," Solid-State Electronics, vol. 14, p. 41, 1971
- [40] P. Calzolari, S. Graffi, and C. Morandi, "Field-Enhanced Carrier Generation in MOS Capacitors," Solid-State Electronics, vol. 17, p. 1001, 1974
- [41] J. Hartke, "The Three-Dimensional Poole-Frenkel Effect," Journal of Applied Physics, vol. 39, p. 4871, 1968
- [42] S. Sze, *Physics of Semiconductor Devices*, 2nd Edition, John Wiley and Sons, 1981
- [43] S. Sze, *Physics of Semiconductor Devices*, 2nd Edition, John Wiley and Sons, p. 103, 1981
- [44] A. Foyt, W. Lindley, C. Wolfe, and J. Donnelly, "Isolation of Junction Devices in GaAs using Proton Bombardment," Solid-State Electronics, vol. 12, p. 209, 1969

- [45] J. Pabst, T. Dungan, J. Cooper, M. Melloch, "Ion-Implanted pn-Junction Capacitors for GaAs DRAMs," submitted for publication, July 1989
- [46] J. Pabst, "Ion-Implanting GaAs for DRAM Applications," Master's Thesis, Purdue University, 1989
- [47] K. Loh, D. Schroeder, R. Clarke, A. Rohatgi, and G. Eldridge, "Low Leakage Current GaAs Diodes," IEEE Transactions on Electron Devices, vol. 28, no. 7, p. 796, 1981
- [48] C. Chang, T. Vrotsos, M. Frizzell, and R. Carroll, "A Subthreshold Current Model for GaAs MESFET's," IEEE Electron Device Letters, vol. 8, no. 2, p. 69, 1987
- [49] J. Conger, A. Peczalski, and M. Shur, "Subthreshold Current in GaAs MESFET's," IEEE Electron Device Letters, vol. 9, no. 3, p. 128, 1988
- [50] J. Marshall and J. Meindl, "An Analytical Two-Dimensional Model for Silicon MESFET's," IEEE Transactions on Electron Devices, vol. 35, no. 3, p. 373, 1988
- [51] H. Bethe, "Theory of the Boundary Layer of Crystal Rectifiers," MIT Radiation Laboratory Report, 43-12, 1942
- [52] W. Schottky, "Halbleiter Theorie der Sperrschicht," Naturwissenschaften, vol. 26, p. 843, 1938
- [53] C. Crowell and S. Sze, "Current Transport in Metal-Semiconductor Barriers," Solid State Electronics, vol. 8, p. 395, 1965
- [54] C. Han, D. Grider, K. Newstrom, P. Joslyn, A. Fraasch, and D. Arch, "0.5 μ m Gate Length Self-Aligned Gate MODFET with Reduced Short Channel Effects," Technical Digest of the 46th Annual Device Research Conference, p. VA-1, 1988
- [55] T. Furuyama, T. Ohsawa, Y. Watanabe, H. Ishiuchi, T. Watanabe, T. Tanaka, K. Natori, and O. Ozawa, "An Experimental 4-Mbit CMOS DRAM," IEEE Journal of Solid State Circuits, vol. 21, no. 5, p. 605, 1986
- [56] B. Van Zeghbroeck, W. Patrick, H. Meier, and P. Vettinger, "Submicrometer GaAs MESFET with Shallow Channel and Very High Transconductance," IEEE Electron Device Letters, vol. 8, no. 3, 1987

- [57] A. Fiedler, J. Chun, and D. Kang, "A 3 Ns 1K X 4 Static Self-Timed GaAs RAM," GaAs IC Symposium Technical Digest, p. 67, 1988
- [58] A. Tasch, Jr., P. Chatterjee, H-S. Fu, and T. Hollaway, "The Hi-C RAM Cell Concept," IEEE Transactions on Electron Devices, vol. 25, p. 33, 1978
- [59] P. Neudeck, T. Dungan, J. Cooper, M. Melloch, "Electrical Characterization of a JFET-Accessed GaAs Dynamic RAM Cell," submitted for publication, July 1989
- [60] P. Neudeck, unpublished results

Appendix A

Journal Publications and Conference Presentations acknowledging support from Contract N00014-86-K-0350 and Theses.

Journal Publications:

- [1] R. L. Gunshor, L. A. Kolodziejski, M. R. Melloch, M. Vaziri, C. Choi, and N. Otsuka, "Nucleation and Characterization of Pseudomorphic ZnSe Grown on Molecular Beam Epitaxially-Grown GaAs Epilayers," *Appl. Phys. Lett.* 50, 200 (1987).
- [2] M. R. Melloch, J. A. Cooper, Jr., and Q-D. Qian, "Investigation of Minority Carrier Retention Behind AlAs Barriers," *Journal of Vac. Sci. and Tech. B* 5, (1987).
- [3] T. E. Dungan, J. A. Cooper, Jr., and M. R. Melloch, "A Thermal-Generation-Limited Buried-Well Structure for Room Temperature GaAs Dynamic RAMs," *IEEE Elec. Dev. Lett.* EDL-8, 243 (1987).
- [4] A. K. Arora, A. K. Ramdas, M. R. Melloch, and N. Otsuka, "Interface vibrational Raman lines in GaAs/ $\text{Al}_x\text{Ga}_{1-x}\text{As}$ superlattices," *Phys. Rev.* B35, (1987).
- [5] M. S. Carpenter, M. R. Melloch, M. S. Lundstrom, S. P. Tobin, "Effects of Na_2S and $(\text{NH}_4)_2\text{S}$ edge passivation treatments on the dark current-voltage characteristics of GaAs pn diodes," *Appl. Phys. Lett.* 52, 2157 (1988).
- [6] M. S. Carpenter, M. R. Melloch, and T. E. Dungan, "Schottky barrier formation on $(\text{NH}_4)_2\text{S}$ -treated n- and p-type (100) GaAs," *Appl. Phys. Lett.* 53, 66 (1988).
- [7] B. A. Cowans, Z. Dardas, W. N. Delgass, M. S. Carpenter, and M. R. Melloch, "X-ray Photoelectron Spectroscopy of Ammonium Sulfide Treated GaAs (100) Surfaces," *Appl. Phys. Lett.* 54, 365 (1989).
- [8] J. S. Kleine, Q.-D. Qian, J. A. Cooper, Jr., and M. R. Melloch, "Electron Emission from Direct Bandgap Heterojunction Capacitors," *IEEE Trans. Elec. Dev.* ED-36, 289 (1989).
- [9] M. R. Melloch, D. C. Miller, and B. Das, "Effect of a GaAs Buffer Layer Grown at Low Substrate Temperatures on a High-Electron-Mobility Modulation-Doped Two-Dimensional Electron Gas," *Appl. Phys. Lett.* 54, 943 (1989).
- [10] Q.-D. Qian, M. R. Melloch, and J. A. Cooper, Jr., "Electrical Behavior of a Static Hole Inversion Layer at the i-AlAs/n-GaAs Heterojunction," *J. Appl. Phys.* 65, 3118 (1989).
- [11] M. S. Carpenter, M. R. Melloch, B. A. Cowans, Z. Dardas, and W. N. Delgass, "Investigations of Ammonium Sulfide Surface Treatments on GaAs," *Journal of Vac. Sci. and Tech.* B7, (1989).
- [12] J. S. Kleine, M. R. Melloch, and J. A. Cooper, Jr., "The Effect of Al mole fraction on Electron Emission at the $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ Interface," submitted to *Applied Physics Letters*.

- [13] J. W. Pabst, T. E. Dungan, J. A. Cooper, Jr., and M. R. Melloch, "Ion-implanted pn-Junction Capacitors for GaAs DRAMS," submitted to IEEE Transactions on Electron Devices.
- [14] P. G. Neudeck, T. E. Dungan, M. R. Melloch, and J. A. Cooper, Jr., "Electrical Characterization of a JFET-Accessed GaAs Dynamic RAM Cell," submitted to IEEE Electron Device Letters.

Conference Presentations:

- [1] M. R. Melloch, J. A. Cooper, Jr., and Q-D. Qian, "Investigation of Minority Carrier Hole Retention Behind AlAs and AlAs/GaAs Superlattice Barriers," Seventh US MBE Workshop, Boston, MA, October 20-22, 1986.
- [2] M. Vaziri, R. L. Gunshor, L. A. Kolodziejski, and M. R. Melloch, "Characterization of ZnSe on GaAs Epilayers," March Meeting of the American Physical Society, New York, New York, March 16-20, 1987.
- [3] L. A. Kolodziejski, R. L. Gunshor, M. Melloch, M. Vaziri, C. Choi, and N. Otsuka, "MBE of ZnSe on GaAs Epilayers," SPIE's conference on Growth of Compound Semiconductors, Bay Point, Florida, March 26-27, 1987.
- [4] G. S. Studtmann, R. L. Gunshor, L. A. Kolodziejski, M. R. Melloch, N. Otsuka, D. P. Munich, J. A. Cooper, and R. F. Pierret, "Pseudomorphic ZnSe/GaAs MISFET Devices," 45th Device Research Conference, University of California at Santa Barbara, June 22-24, 1987.
- [5] M. S. Carpenter, M. R. Melloch, and T. E. Dungan, "Formation of Schottky Barriers on Reduced Surface State GaAs," 1988 International Symposium on Gallium Arsenide and Related Compounds, Atlanta, Georgia, September 11-14, 1988.
- [6] J. S. Kleine, Q-D Qian, J. A. Cooper, Jr., and M. R. Melloch, "Mathematical Model for Carrier Emission from a Static Two-Dimensional Electron Gas," 1988 International Symposium on Gallium Arsenide and Related Compounds, Atlanta, Georgia, September 11-14, 1988.
- [7] M. S. Carpenter, M. R. Melloch, B. A. Cowans, and W. N. Delgass, "Investigations of $(\text{NH}_4)_2\text{S}$ Surface Treatments on GaAs," 16th Conference on Physics and Chemistry of Semiconductor Interfaces, Bozeman, Montana, Feb. 7-9, 1989.

Theses:

- [1] John W. Pabst, "Ion-Implanting GaAs for DRAM Applications," MSEE thesis, August 1989.

- [2] Thomas E. Dungan, "Dynamic Memories for Gallium Arsenide," Ph.D. thesis, August 1989.